

IN THE UNITED STATES DISTRICT COURT
FOR THE SOUTHERN DISTRICT OF TEXAS
CORPUS CHRISTI DIVISION

MARC VEASEY, *et al.*,

Plaintiffs,

v.

RICK PERRY, *et al.*,

Defendants.

Civil Action No. 2:13-cv-193 (NGR)

UNITED STATES OF AMERICA,

Plaintiff,

TEXAS LEAGUE OF YOUNG VOTERS
EDUCATION FUND, *et al.*,

Plaintiff-Intervenors,

TEXAS ASSOCIATION OF HISPANIC
COUNTY JUDGES AND COUNTY
COMMISSIONERS, *et al.*,

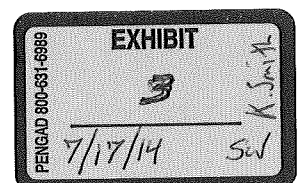
Plaintiff-Intervenors,

v.

STATE OF TEXAS, *et al.*,

Defendants.

Civil Action No. 2:13-cv-263 (NGR)



TEXAS STATE CONFERENCE OF NAACP
BRANCHES, *et al.*,

Plaintiffs,

v.

JOHN STEEN, *et al.*,

Defendants.

Civil Action No. 2:13-cv-291 (NGR)

BELINDA ORTIZ, *et al.*,

Plaintiffs,

v.

STATE OF TEXAS, *et al.*,

Defendants

Civil Action No. 2:13-cv-348 (NGR)

DECLARATION OF KENNETH SMITH

I, Kenneth Smith, pursuant to 28 U.S.C. § 1746, declare as follows:

1. I currently serve as the Assistant Director in the Veterans Benefits Administration's (VBA) Office of Performance Analysis and Integrity. I have served in this capacity for three years, nine months. My duties include the collection and reporting of VBA business and operational data sourced from the Enterprise Data Warehouse.
2. I have personal knowledge of the information contained in this declaration based upon my work for the Veterans Benefits Administration ("VBA") at the U.S. Department of Veterans Affairs ("VA"), and my supervision of the completion of the database comparison described below.

3. On March 5, 2014, the VA received from the United States Department of Justice an encrypted hard drive containing data regarding Texas registered voters in the form of a 7 gigabyte .csv file containing 13,564,420 records for Texas registered voters. VBA was able to load 13,555,657 of the 13,564,420 records. The 8,763 voter records that VBA was not able to load and analyze as part of the database comparison process (0.0006% of the Texas voter registration records received) contained embedded commas in an address or name field. These embedded commas prevented VBA from being able to load and analyze only those affected records.

4. The Texas data that VBA received contained the following fields:

- | | |
|---------------------|-------------------|
| - Vuid | - str_num_res_alt |
| - last_name | - AGDN |
| - first_name | - AGDN_TAG |
| - middle_name | - AGDN_unique |
| - ssn | - AGDNlast |
| - ssn4_str | - AGDNlast_TAG |
| - gender_code | - AGDNlast_unique |
| - official_id | - AGD |
| - dl_id_str | - AGD_TAG |
| - date_of_birth | - AGD_unique |
| - voter_status_code | - and |
| - block_number | - ADN_TAG |
| - street_name | - ADN_unique |
| - city | - AGN |
| - state | - AGN_TAG |
| - zip_code | - AGN_unique |
| - zip_str | - GDN |
| - mail_adrs_1 | - GDN_TAG |
| - last_fix | - GDN_unique |
| - multi_lastname1 | - DLN |
| - multi_lastname2 | - DLN_TAG |
| - first_fix | - DLN_unique |
| - middle_initial | - unique_count |
| - fem_str | - FMLD |
| - dob_str | - FMLD_TAG |
| - zip5_res_alt | - FMLD_unique |

- | | |
|----------------|-----------------|
| - SDZ | - FML2D |
| - SDZ_TAG | - FML2D_TAG |
| - SDZ_unique | - FML2D_unique |
| - SND | - unique_count2 |
| - SND_TAG | - var_fill |
| - SND_unique | - K2 |
| - FML1D | - K3 |
| - FML1D_TAG | - L2 |
| - FML1D_unique | - L3 |

5. VBA executed the data preparation and comparison steps, attached as Ex. A, that were provided by the United States Department of Justice.
6. Through this process, VBA attempted to match particular combinations of identifying information for Texas registered voters (*e.g.*, first and last name, gender, and date of birth) with the same or related combinations of identifying information with respect to records for living Veterans with a 50% or higher disability rating, either service connected or non-service connected. For each particular combination, VBA identified instances where the identifying information for a Texas registered voter matched with the respective combination of identifying information for one or more living Veterans with a disability rating of 50% or higher.
7. For each sweep through the relevant VBA datasets, we appended a column to the Texas data to indicate, on a record-by-record basis, the output of the database comparisons on all of the requested combinations. The frequencies of any missing values in the underlying VBA data are provided as Ex. B.
8. On May 5, 2014, VBA transferred responsive data to the Department of Justice, in the form of a 3.2 gigabyte, zipped variable width file with pipe delimiters. The unzipped file is 37.8 gigabytes. The name of the results file transferred to the Department of Justice is tx_folder_project.texas_results_table_final_output.

9. The results of each sweep, as described in Ex. A, can be found in the following columns of the data that VBA returned to the Department of Justice on May 5:

- a. "USA_MATCHES_ID" contains results of the Stage 1, Step 3.1.1, Combination A match;
- b. "USB_MATCHES_ID" contains results of the Stage 1, Step 3.1.2, Combination B match;
- c. "USC_MATCHES_ID" contains results of the Stage 1, Step 3.1.2, Combination C match;
- d. "USD_MATCHES_ID" contains results of the Stage 1, Step 3.1.2, Combination D match;
- e. "USE1_MATCHES_ID" contains results of the Stage 1, Step 3.1.2, Combination E match;
- f. "USF_MATCHES_ID" contains results of the Stage 1, Step 3.1.2, Combination F match;
- g. "USG_MATCHES_ID" contains results of the Stage 2, Step 3.2.1, Combination G match;
- h. "USH_MATCHES_ID" contains results of the Stage 2, Step 3.2.2, Combination H match;
- i. "USI_MATCHES_ID" contains results of the Stage 2, Step 3.2.2, Combination I match;
- j. "USSSN_MATCHES_ID" contains results of the Stage 2, Step 3.2.2, nine-digit social security number match;

- k. "USK2_MATCHES_ID" contains results of the Stage 2, Step 3.2.3, Combination K to G match;
- l. "USK_MATCHES_ID" contains results of the Stage 2, Step 3.2.3, Combination K match;
- m. "USK3_MATCHES_ID" contains results of the Stage 2, Step 3.2.3, Combination K to L match;
- n. "USL2_MATCHES_ID" contains results of the Stage 2, Step 3.2.4, Combination L to G match;
- o. "USL_MATCHES_ID" contains results of the Stage 2, Step 3.2.4, Combination L match;
- p. "USL3_MATCHES_ID" contains results of the Stage 2, Step 3.2.4, Combination L to K match;
- q. "USF_NW_MATCHES_ID" contains results of the Stage 3, Step 3.3.1, Combination F match;
- r. "USG_NW_MATCHES_ID" contains results of the Stage 3, Step 3.3.2, Combination G match;
- s. "USI_NW_MATCHES_ID" contains results of the Stage 3, Step 3.3.2, Combination I match;
- t. "USSSN_NW_MATCHES_ID" contains results of the Stage 3, Step 3.3.2, nine-digit social security number match;
- u. "USK2_NW_MATCHES_ID" contains results of the Stage 3, Step 3.3.3, Combination K to G match;

- v. "USK_NW_MATCHES_ID" contains results of the Stage 2, Step 3.3.3, Combination K match;
- w. "USK3_NW_MATCHES_ID" contains results of the Stage 2, Step 3.3.3, Combination K to L match;
- x. "USL2_NW_MATCHES_ID" contains results of the Stage 3, Step 3.3.4, Combination L to G match;
- y. "USL_NW_MATCHES_ID" contains results of the Stage 3, Step 3.3.4, Combination L match;
- z. "USL3_NW_MATCHES_ID" contains results of the Stage 3, Step 3.3.4, Combination L to K match;
- aa. "TEXAS1_MATCHES_ID" contains results of the Stage 4, Step 3.4.1, Sweep 1 match;
- bb. "TEXAS2_MATCHES_ID" contains results of the Stage 4, Step 3.4.2, Sweep 2 match;
- cc. "TEXAS3_MATCHES_ID" contains results of the Stage 4, Step 3.4.3, Sweep 3 match; and
- dd. "TEXAS4_MATCHES_ID" contains results of the Stage 4, Step 3.4.4, Sweep 4 match.

10. No other matching comparisons between the Texas data and data from the VBA were undertaken beyond those set forth above, in accordance with the steps provided in Ex. A.

I declare under penalty of perjury that the foregoing is true and correct.

Executed on May 27, 2014.




Exhibit A

United States v. Texas: Federal Agency Algorithm Summary

This document summarizes the database matches that both the United States and all other plaintiffs and the State of Texas have requested from Federal Agencies as part of the *Veasey v. Perry/United States v. Texas* litigation (S.D. Tex).

The matching process proceeds in three parts, which are explained in detail below. *First*, databases are prepared and standardized. *Second*, identifier values are constructed by combining multiple individual fields. *Third*, the United States' one-to-many matches and the State of Texas's many-to-many matches are conducted between databases.

PART I: DATABASE PREPARATION

Stage 1: Extraction of Available Data from Federal Identification & Disability Databases

- Step 1.1.1:** Extract complete name into separate first name, middle name, and last name fields.
- Step 1.1.2:** Extract date of birth.
- Step 1.1.3:** Extract gender.
- Step 1.1.4:** Extract residential address and mailing address.
- Step 1.1.5:** Extract social security number.
- Step 1.1.6:** Extract Texas driver license number (only if present in Federal database).

Stage 2: Separate Valid Identification and Disability Records

- Step 1.2.1:** Remove records from identification database extracts that indicate that an ID has been revoked or has expired more than 60 days before the date of the TEAM database snapshot (which is January 15, 2014).

Exhibit A

- Step 1.2.2:** Remove records from disability database extracts that do not indicate current disability status or indicate a Veterans Administration disability rating of less than 50%.

Stage 3: Diagnostics

- Step 1.3.1:** Report the frequency of missing values for each field.
- Step 1.3.2:** Report the frequencies of invalid Social Security numbers, such as 111111111 and 123456789.
- Step 1.3.3:** Report the frequencies of likely invalid dates of birth, such as January 1, 1901 and November 11, 1911.

Stage 4: Standardize Last Name

- Step 1.4.1:** Remove last name suffixes that are contained within the last name field, rather than a distinct suffix field. *E.g.*, <Smith Jr.> becomes <Smith>.
- Step 1.4.2:** For last names containing hyphens, populate separate last name fields for all parts of the last name. *E.g.*, the last name <Smith-Jones> would have the value <Smith> entered into a LastName1 field and the value <Jones> entered into a LastName2 field.
- Step 1.4.3:** Remove spaces, hyphens, periods, and apostrophes from all last name fields and convert all letters to uppercase. *E.g.*, <O'Connor> becomes <OCONNOR> and <Smith-Jones> becomes <SMITHJONES>.
- Step 1.4.4:** Code all missing values as blank fields.

Exhibit A

Stage 5: Standardize First Name and Middle Name

- Step 1.5.1:** Remove spaces, hyphens, periods, and apostrophes from the first name field and convert all letters to uppercase. *E.g.*, <Jean-Paul> becomes <JEANPAUL>.
- Step 1.5.2:** Parse the first letter of the middle name (if available) and use it to populate a middle initial field. *E.g.*, <John> would yield <J>.¹
- Step 1.5.3:** Code all missing values as blank fields.

Stage 6: Standardize Date of Birth

- Step 1.6.1:** Convert the date of birth to an eight-digit string of MMDDYYYY.
- Step 1.6.2:** Code all missing values as blank fields.

Stage 7: Standardize Gender

- Step 1.7.1:** Code gender as a string of 1 for females and 0 for males.
- Step 1.7.2:** Fill missing gender values using the most common gender value for the first name associated with a record. *E.g.*, if 99% of records with first name <JOHN> are listed as male, assign the male identifier to all records with first name <JOHN> and no listed gender.
- Step 1.7.3:** If missing values remain, code all missing values as blank fields.

¹ The U.S. Department of State does not maintain a separate field for middle names in its database of U.S. Passport and Passport Card holders. Instead, both first and middle name may be stored in the first name field. For this database, the following rule will be applied: treat the first word in the first name field as the first name, and treat the first letter following the first space as the middle initial.

Exhibit A

Stage 8: Standardize Address

- Step 1.8.1:** Convert the residential ZIP code to a string if it is stored as a numeric field.
- Step 1.8.2:** Where the residential address ZIP code is blank, populate that field with the value in the mailing address ZIP code field, if available.²
- Step 1.8.3:** Truncate the residential ZIP code field to the first five digits. *E.g.*, <77777-1234> becomes <77777>.
- Step 1.8.4:** Where the residential address field is blank, populate that field with the value in the mailing address field, if available.
- Step 1.8.5:** Where address field containing street address begins with a street number, isolate the street number. *E.g.*, <123 Main Street> becomes <123>.
- Step 1.8.6:** Where the address field begins with recognized strings indicating a mail box, eliminate strings to isolate the box number. *E.g.*, <PO Box 444> becomes <444>.
- Step 1.8.7:** If missing values remain, code all missing values as blank fields.

Stage 9: Standardize Social Security Number

- Step 1.9.1:** Convert the social security number to a string if it is stored as a numeric field.
- Step 1.9.2:** Using full social security number, check for invalid SSNs. In the case of invalid SSNs, code as missing. *E.g.*, <123456789> becomes <>.

² For purposes of this database matching protocol, the only address fields utilized with respect to data regarding U.S. Passports and U.S. Passport Cards are those regarding mailing addresses.

Exhibit A

Step 1.9.3: Extract the last four digits of full social security number as a four-character string and use them to populate a separate SSN4 field.

Step 1.9.4: Code all missing values as blank fields.

PART II: DATABASE PREPARATION

Stage 1: Construct Primary Identifier Variables for United States' One-to-Many Sweeps

Step 2.1.1: Create Combination A: First Name + Last Name + Gender + DOB + Residential ZIP + Residential Street Number. *E.g.*, the separate fields <JEAN>, <SMITH>, <0>, <01011950>, <77777>, and <123> are combined to a single field <JEANSMITH00101195077777123>.³

Step 2.1.2: Create Combination B: Last Name + Gender + DOB + Residential ZIP + Residential Street Number.

Step 2.1.3: Create Combination C: Gender + Date of Birth + Residential ZIP + Residential Street Number.

Step 2.1.4: Create Combination D: First Name + Last Name + Date of Birth + Residential ZIP + Residential Street Number.

Step 2.1.5: Create Combination E: First Name + Last Name + Gender + Residential ZIP + Residential Street Number.

Step 2.1.6: Create Combination F: First Name + Last Name + Gender + DOB.

Step 2.1.7: Create Combination M: Texas Driver License Number (where available).

³ For the U.S. Department of State only, the name portion of any combination is truncated if it is more than 32 characters long.

Exhibit A

Stage 2: Construct Secondary Identifier Variables for United States' One-to-Many Sweeps

Step 2.2.1: Create Combination G: First Name + Middle Initial + Last Name + Date of Birth.⁴

Step 2.2.2: Create Combination H: SSN4 + Date of Birth + Residential ZIP.

Step 2.2.3: Create Combination I: SSN4 + First Name + Last Name + Date of Birth.

Step 2.2.4: Create Combination K: First Name + Last Name 1 + Middle Initial + Date of Birth.

Step 2.2.5: Create Combination L: First Name + Last Name 2 + Middle Initial + Date of Birth.

Step 2.2.6: Full Social Security Number.

Stage 3: Construct Identifiers Used Only For Texas's Many-to-Many Sweeps

Step 2.3.1: Create Combination for Texas's Sweep 1: SSN4 + Last Name + DOB.

Step 2.3.2: Create Combination for Texas's Sweep 3: First Name + Last Name + DOB

Step 2.3.3: Create Combination for Texas's Sweep 4: First Name + Middle Initial + Last Name + DOB

Note: Combinations for Texas's Sweeps 1 and 3 do not already exist as pre-made fields in the TEAM database extract but instead must be created from the underlying TEAM database fields, in addition to being constructed on the Federal database side. Texas's Sweep 4 is equivalent to

⁴ Only for the State Department, create three further variations of Combination G created using the State Department's "LFMName" field which contains Last, First, and Middle Names, in that order, truncated to a maximum length of 32 characters. Combination G1 is DOB + LFMName; Combination G2 is DOB + First two words of LFMName; and Combination G3 is DOB + First two words of LFMName + First character of third word of LFMName.

Exhibit A

the combination for the United States' Combination G. Texas's Sweep 2 is on full 9 social security number.⁵

Stage 4: Establish Identifier Uniqueness For Combinations A - L

Step 2.4.1: Generate a field that establishes the uniqueness of each identifier variable.

For federal databases, for each combination A-L, generate a field that establishes uniqueness among only Texas records and a field that establishes uniqueness among nationwide records. *E.g.*, if only one record has the string <JEANSMITH01011950012377777> for Combination A, populate the uniqueness field for Combination A for that record as <1>. If four records have the string <JOHNSMITHA0101950> for Combination G, populate the uniqueness field for Combination G for each of those records as <2>, which indicates any number greater than one.

⁵ For purposes of matching to U.S. Department of State Passport and Passport Card holder data, Texas's Sweep 3 and Sweep 4 are as follows: Sweep 3: Last Name + First Name (restricted to 32 characters) + DOB. Sweep 4: Last Name + First Name + Middle Initial (restricted to 32 characters) + DOB.

Exhibit A

PART III: MATCH DATABASES

Stage 1: United States' Primary One-to-Many Matching Sweeps

- Step 3.1.1:** For each case in which Combination A is unique in the TEAM database, match Combination A against Combination A in the identifier or disability database. For federal databases, use only the subset of records with Texas addresses in the identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination A output field. Where there is a match, indicate the uniqueness of Combination A in the identifier or disability database in the Combination A output field (*e.g.*, in cases where there is one matching record in the Federal database, <1> should be inserted into the Combination A output field, while a <2> should be inserted into the Combination A output field if the TEAM record matched 2 or more records in the Federal database).
- Step 3.1.2:** Use the procedure in Step 3.1.1 to match Combination B, Combination C, Combination D, Combination E, and Combination F in the TEAM database against the equivalent combination field in the identifier or disability database.
- Step 3.1.3:** Use the procedure in Step 3.1.1 to match Combination M in the TEAM database against the equivalent combination field in the identifier databases produced by the State of Texas.

Exhibit A

Stage 2: United States' Secondary One-to-Many Matching Sweeps

- Step 3.2.1:** For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), and where Combination G is unique in the TEAM database, match Combination G against Combination G in the identifier or disability database. For federal databases, use only the subset of records with Texas addresses in the identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination G output field. Where there is a match, indicate the uniqueness of Combination G in the identifier or disability database in the Combination G output field (*e.g.*, <1> if a unique match and <2> if matched to more than one record).
- Step 3.2.2:** For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination H, Combination I, and complete social security number⁶ in the TEAM database against the equivalent combination/field in the identifier or disability database.
- Step 3.2.3:** For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination K against Combination G, Combination K, and Combination L in the identifier or disability database.

⁶ The full social security number is not created as a separate “combination” as it is its own field stored within the TEAM database under the field name “ssn”.

Exhibit A

Step 3.2.4: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination L against Combination G, Combination K, and Combination L in the identifier or disability database.

Stage 3: United States' Nationwide Federal Sweeps

Step 3.3.1: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, and where Combination F is unique, match Combination F against Combination F in the nationwide identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination F nationwide output field. Where there is a match, indicate the uniqueness of Combination F in the identifier or disability database in the Combination F nationwide output field (*e.g.*, <1> if a unique match and <2> if matched to more than one record).

Step 3.3.2: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination G, Combination I, and full social security number in the TEAM database against the equivalent combination/field in the nationwide identifier or disability database.

Step 3.3.3: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination

Exhibit A

K against Combination G, Combination K, and Combination L in the nationwide identifier or disability database.

Step 3.3.4: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination L against Combination G, Combination K, and Combination L in the nationwide identifier or disability database.⁷

Stage 4: Texas' Many-to-Many Nationwide Sweeps

Step 3.4.1 Regardless of whether the combination for Sweep 1 is unique in the TEAM database, match against the equivalent combination in a nationwide search of the Federal database.

⁷ Step 3.3.5 for the State Department only: Match the following Combination G variations from applicable State Department records, first to include only the subset of records with Texas addresses, and then to include all applicable U.S. Passport and Passport Card records nationwide (*e.g.*, without Texas addresses), against the following fields from the TEAM database:

- Combination G1 to DOB + Last_fix + First_fix + Middle_name from the TEAM database;
- Combination G2 to DOB + Last_fix + First_fix from the TEAM database;
- Combination G2 to DOB + Last_fix + First word of First_name from the TEAM database;
- Combination G3 to DOB + Last_fix + First_fix + Middle_Initial from the TEAM database;
- Combination G3 to DOB + Last_fix + First word of First_name + Middle_Initial from the TEAM database; and
- Combination G3 to DOB + Last_fix + First word of First_name + First character of Second word of First_fix from the TEAM database.

Attempt matches for all TEAM records, regardless of whether they matched in any prior sweeps. Indicate <1> if a unique match and <2> if matched to more than one record.

Exhibit A

Step 3.4.2 Regardless of whether full 9 social security number is unique in the TEAM database, for Sweep 2, match against the equivalent field in a nationwide search of the Federal database.

Step 3.4.3 Regardless of whether the combination for Sweep 3 is unique in the TEAM database, match against the equivalent combination in a nationwide search of the Federal database.

Step 3.4.4 Regardless of whether the combination for Sweep 4 (Combination G) is unique in the TEAM database, match against the equivalent combination in a nationwide search of the Federal database.

Note: For each of the Texas many-to-many sweeps:

- Indicate <1> if any TEAM combination matches a single combination in the Federal database
- Indicate <2> if any TEAM combination matches more than one record in the Federal database.
- Indicate <0> if no match is achieved.

Examples:

- If there are two TEAM records that have identical versions of the combination for Sweep 1, and there is one record in the Federal database that matches on this combination, both of the underlying TEAM records will have a matching output of <1> for Sweep 1.
- If there are three TEAM records that have identical versions of the combination for Sweep 3, and there are five records in the Federal database that match on that combination, the three TEAM records will each have a matching output of <2> for Sweep 3.

Exhibit B

Veterans Benefits Administration

Step 1.3 Diagnostics on VBA Data

Total number of VBA records for all living Veterans rated as having 50% or higher disability, either service connected or non-service connected: 2,203,046

Blank field counts:

Last Name – 0

First Name – 0

Middle Name – 261,9448, or 11.89%

SSN – 4,211, or 0.191%

Date of Birth – 316, or 0.014%

Gender – 41,812, or 1.898%

Mailing Address – 1,639 or 0.074%

Residence Address – 1,735,958, or 78.798%

Counts of invalid SSNs:

111111111 – 0

222222222 – 0

333333333 – 0

444444444 – 0

555555555 – 0

666666666 – 0

777777777 – 0

888888888 – 0

999999999 – 0

000000000 – 0

012345678 – 0

123456789 – 0

987654321 – 0

Counts of potentially invalid dates of birth:

01-01-1901 – 5

02-02-1902 – 0

03-03-1903 – 0

04-04-1904 – 0

05-05-1905 – 0

06-06-1906 – 0

07-07-1907 – 1

08-08-1908 – 0

09-09-1909 – 0

10-10-1910 – 0

11-11-1911 – 1

12-12-1912 – 1

IN THE UNITED STATES DISTRICT COURT
FOR THE SOUTHERN DISTRICT OF TEXAS
CORPUS CHRISTI DIVISION

MARC VEASEY, *et al.*,

Plaintiffs,

v.

RICK PERRY, *et al.*,

Defendants.

Civil Action No. 2:13-cv-193 (NGR)

UNITED STATES OF AMERICA,

Plaintiff,

TEXAS LEAGUE OF YOUNG VOTERS
EDUCATION FUND, *et al.*,

Plaintiff-Intervenors,

TEXAS ASSOCIATION OF HISPANIC
COUNTY JUDGES AND COUNTY
COMMISSIONERS, *et al.*,

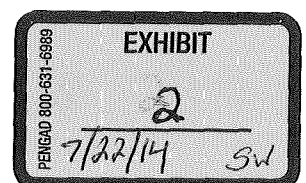
Plaintiff-Intervenors,

v.

STATE OF TEXAS, *et al.*,

Defendants.

Civil Action No. 2:13-cv-263 (NGR)



TEXAS STATE CONFERENCE OF NAACP
BRANCHES, *et al.*,

Plaintiffs,

v.

JOHN STEEN, *et al.*,

Defendants.

Civil Action No. 2:13-cv-291 (NGR)

BELINDA ORTIZ, *et al.*,

Plaintiffs,

v.

STATE OF TEXAS, *et al.*,

Defendants

Civil Action No. 2:13-cv-348 (NGR)

DECLARATION OF LEE CHARLES BAYDUSH

I, Lee Charles Baydush, pursuant to 28 U.S.C. § 1746, declare as follows:

1. I currently serve as subject matter expert for the U.S. Department of State, Bureau of Consular Affairs, Consular Systems & Technology, Service integration & Innovation Division (CA/CST/SII). I have served in this capacity for eleven months. My duties include research, analysis, consulting, and operational support of Consular IT systems.
2. I have personal knowledge of the information contained in this declaration based upon my work for the State Department, and my personal involvement in completion of the database comparison described below.
3. On February 27, 2014, the State Department received an encrypted hard drive containing data regarding Texas registered voters from the United States Department of Justice in

the form of a 7 gigabyte .csv file containing 13,564,420 records for Texas registered voters. The State Department was able to load all records except for 6. The identifiers (VUIDs) for 5 of the 6 records that my agency was not able to load and analyze as part of the database comparison process are as follows: 1018504714; 1018510162; 1018555565; 1111894862; 1164253546. The final record that could not be loaded lacked a VUID, but has a unique identifier of 3890395vf (other than this unique identifier “var_fill” value, this record was blank).

4. The Texas data that the State Department received contained the following fields:

- | | |
|---------------------|-------------------|
| - Vuid | - AGDN_TAG |
| - last_name | - AGDN_unique |
| - first_name | - AGDNlast |
| - middle_name | - AGDNlast_TAG |
| - ssn | - AGDNlast_unique |
| - ssn4_str | - AGD |
| - gender_code | - AGD_TAG |
| - official_id | - AGD_unique |
| - dl_id_str | - and |
| - date_of_birth | - ADN_TAG |
| - voter_status_code | - ADN_unique |
| - block_number | - AGN |
| - street_name | - AGN_TAG |
| - city | - AGN_unique |
| - state | - GDN |
| - zip_code | - GDN_TAG |
| - zip_str | - GDN_unique |
| - mail_adrs_1 | - DLN |
| - last_fix | - DLN_TAG |
| - multi_lastname1 | - DLN_unique |
| - multi_lastname2 | - unique_count |
| - first_fix | - FMLD |
| - middle_initial | - FMLD_TAG |
| - fem_str | - FMLD_unique |
| - dob_str | - SDZ |
| - zip5_res_alt | - SDZ_TAG |
| - str_num_res_alt | - SDZ_unique |
| - AGDN | - SND |

- | | |
|----------------|-----------------|
| - SND_TAG | - FML2D_unique |
| - SND_unique | - unique_count2 |
| - FML1D | - var_fill |
| - FML1D_TAG | - K2 |
| - FML1D_unique | - K3 |
| - FML2D | - L2 |
| - FML2D_TAG | - L3 |

5. The State Department executed the data preparation and comparison steps, attached as Ex. A, that were provided by the United States Department of Justice.
6. Through this process, the State Department attempted to match particular combinations of identifying information for Texas registered voters (*e.g.*, first and last name, gender, and date of birth) with the same or related combinations of identifying information with respect to holders of U.S. Passports and U.S. Passport cards. For each particular combination, the State Department identified instances where the identifying information for a Texas registered voter matched with the respective combination of identifying information for one or more holders of a U.S. Passport or U.S. Passport Card.
7. For each sweep through the relevant State Department datasets, we appended a column to the Texas data to indicate, on a record-by-record basis, the output of the database comparisons on all of the requested combinations.
8. Two State Department datasets were created for the analysis—one contained qualifying records of U.S. Passport and U.S. Passport Card holders with a Texas mailing address and the other contained qualifying records of U.S. Passport and U.S. Passport Card holders without a Texas mailing address. The frequencies of any missing values in those datasets are provided as Ex. B.
9. On April 9, 2014, the State Department transferred responsive data to the Department of Justice, in the form of a variable width file with comma delimiters. In this file, there

were 7,741 records in the TEAM database with commas embedded in the address field, and several hundred records with commas embedded in various name fields. Such commas could be incorrectly treated as delimiters in reading the Department of State matching results. As a result, I re-exported the results file using pipe delimiters and on May 14, 2014, the State Department transferred to the Department of Justice a compressed and encrypted 1.7 gigabyte zip file on a DVD that contained a 7.44 gigabyte pipe-delimited file. Both the April 9 and May 14 files are identical other than the use of different delimiters. The use of pipe delimiters on the May 14, 2014, file will prevent any difficulties reading the embedded commas in the results file. The name of the May 14, 2014 results file is Team_Pipe_Delim.

10. The results of each sweep, as described in Ex. A, can be found in the following columns of the data that the State Department returned to the Department of Justice on May 14:
 - a. "AGDN_matches_id" contains results of the Stage 1, Step 3.1.1, Combination A match;
 - b. "AGDNlast_matches_id" contains results of the Stage 1, Step 3.1.2, Combination B match;
 - c. "AGD_matches_id" contains results of the Stage 1, Step 3.1.2, Combination C match;
 - d. "ADN_matches_id" contains results of the Stage 1, Step 3.1.2, Combination D match;
 - e. "AGN_matches_id" contains results of the Stage 1, Step 3.1.2, Combination E match;
 - f. "GDN_matches_id" contains results of the Stage 1, Step 3.1.2, Combination F match;
 - g. "FMLD_matches_id" contains results of the Stage 2, Step 3.2.1, Combination G match;

- h. "SDZ_matches_id" contains results of the Stage 2, Step 3.2.2, Combination H match;
- i. "SND_matches_id" contains results of the Stage 2, Step 3.2.2, Combination I match;
- j. "ssn_matches_id" contains results of the Stage 2, Step 3.2.2, nine-digit social security number match;
- k. "FML1D_matches_id_Other_FMLD" contains results of the Stage 2, Step 3.2.3, Combination K to G match;
- l. "FML1D_matches_id_Other_FML1D" contains results of the Stage 2, Step 3.2.3, Combination K match;
- m. "FML1D_matches_id_Other_FML2D" contains results of the Stage 2, Step 3.2.3, Combination K to L match;
- n. "FML2D_matches_id_Other_FMLD" contains results of the Stage 2, Step 3.2.4, Combination L to G match;
- o. "FML2D_matches_id_Other_FML2D" contains results of the Stage 2, Step 3.2.4, Combination L match;
- p. "FML2D_matches_id_Other_FML1D" contains results of the Stage 2, Step 3.2.4, Combination L to K match;
- q. "GDN_matches_id_Other_GDN" contains results of the Stage 3, Step 3.3.1, Combination F match;
- r. "FMLD_matches_id_Other_FMLD" contains results of the Stage 3, Step 3.3.2, Combination G match;
- s. "SND_matches_id_Other_SND" contains results of the Stage 3, Step 3.3.2, Combination I match;

- t. "ssn_matches_id_Other" contains results of the Stage 3, Step 3.3.2, nine-digit social security number match;
- u. "FML1D_matches_id_Other_FMLD" contains results of the Stage 3, Step 3.3.3, Combination K to G match;
- v. "FML1D_matches_id_Other_FML1D" contains results of the Stage 2, Step 3.3.3, Combination K match;
- w. "FML1D_matches_id_Other_FML2D" contains results of the Stage 2, Step 3.3.3, Combination K to L match;
- x. "FML2D_matches_id_Other_FMLD" contains results of the Stage 3, Step 3.3.4, Combination L to G match;
- y. "FML2D_matches_id_Other_FML2D" contains results of the Stage 3, Step 3.3.4, Combination L match;
- z. "FML2D_matches_id_Other_FML1D" contains results of the Stage 3, Step 3.3.4, Combination L to K match;
- aa. "a1_matches_id_TX" and "a1_matches_id_Other" respectively contain the results of the Stage 3, Step 3.3.5 Combination G1 to DOB + Last_fix + First_fix + Middle_name matches for the Texas-only and non-Texas address datasets;
- bb. "a2_1_matches_id_TX" and "a2_1_matches_id_Other" respectively contain the results of the Stage 3, Step 3.3.5 Combination G2 to DOB + Last_fix + First_fix matches for the Texas-only and non-Texas address datasets;
- cc. "a2_2_matches_id_TX" and "a2_2_matches_id_Other" respectively contain the results of the Stage 3, Step 3.3.5 Combination G2 to DOB + Last_fix + First word of First_name matches for the Texas-only and non-Texas address datasets;

- dd. "a3_1_matches_id_TX" and "a3_1_matches_id_Other" respectively contain the results of the Stage 3, Step 3.3.5 Combination G3 to DOB + Last_fix + First_fix + Middle_Initial matches for the Texas-only and non-Texas address datasets;
- ee. "a3_2_matches_id_TX" and "a3_2_matches_id_Other" respectively contain the results of the Stage 3, Step 3.3.5 Combination G3 to DOB Last_fix + First word of First_name + Middle_Initial matches for the Texas-only and non-Texas address datasets;
- ff. "a3_3_matches_id_TX" and "a3_3_matches_id_Other" respectively contain the results of the Stage 3, Step 3.3.5 Combination G3 to DOB Last_fix + First word of First_name + First character of Second word of First_fix matches for the Texas-only and non-Texas address datasets;
- gg. "TXMM_snlastdAll_matches_id" contains results of the Stage 4, Step 3.4.1, Sweep 1 match, while "TXMM_TX_snlastdAll_matches_id" contains the results for the Sweep 1 match based on the Texas-only dataset and "TXMM_Other_snlastdAll_matches_id" contains the results for the Sweep 1 match based on the non-Texas addresses dataset;
- hh. "TXMM_ssn_matches_id" contains results of the Stage 4, Step 3.4.2, Sweep 2 match, while "TXMM_TX_ssn_matches_id" contains the results for the Sweep 2 match based on the Texas-only dataset and "TXMM_Other_ssn_matches_id" contains the results for the Sweep 2 match based on the non-Texas address dataset;
- ii. "TXMM_ndAll_matches_id" contains results of the Stage 4, Step 3.4.3, Sweep 3 match, while "TXMM_TX_ndAll_matches_id" contains the results for the Sweep 3 match based on the Texas-only dataset and "TXMM_Other_ndAll_matches_id"

contains the results for the Sweep 3 match based on the non-Texas address dataset;
and

jj. "TXMM_fmldAll_matches_id" contains results of the Stage 4, Step 3.4.4, Sweep 4 match, while "TXMM_TX_fmldAll_matches_id" contains the results for the Sweep 4 match based on the Texas-only dataset and "TXMM_Other_fmldAll_matches_id" contains the results for the Sweep 4 match based on the non-Texas address dataset.

11. No other matching comparisons between the Texas data and data from State Department were undertaken beyond those set forth above, in accordance with the steps provided in Ex. A.

I declare under penalty of perjury that the foregoing is true and correct.

Executed on May 21, 2014.

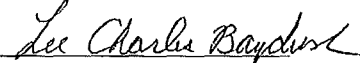

Lee Charles Baydus

Exhibit A

United States v. Texas: Federal Agency Algorithm Summary

This document summarizes the database matches that both the United States and all other plaintiffs and the State of Texas have requested from Federal Agencies as part of the *Veasey v. Perry/United States v. Texas* litigation (S.D. Tex).

The matching process proceeds in three parts, which are explained in detail below. *First*, databases are prepared and standardized. *Second*, identifier values are constructed by combining multiple individual fields. *Third*, the United States' one-to-many matches and the State of Texas's many-to-many matches are conducted between databases.

PART I: DATABASE PREPARATION

Stage 1: Extraction of Available Data from Federal Identification & Disability Databases

- Step 1.1.1:** Extract complete name into separate first name, middle name, and last name fields.
- Step 1.1.2:** Extract date of birth.
- Step 1.1.3:** Extract gender.
- Step 1.1.4:** Extract residential address and mailing address.
- Step 1.1.5:** Extract social security number.
- Step 1.1.6:** Extract Texas driver license number (only if present in Federal database).

Stage 2: Separate Valid Identification and Disability Records

- Step 1.2.1:** Remove records from identification database extracts that indicate that an ID has been revoked or has expired more than 60 days before the date of the TEAM database snapshot (which is January 15, 2014).

Exhibit A

- Step 1.2.2:** Remove records from disability database extracts that do not indicate current disability status or indicate a Veterans Administration disability rating of less than 50%.

Stage 3: Diagnostics

- Step 1.3.1:** Report the frequency of missing values for each field.
- Step 1.3.2:** Report the frequencies of invalid Social Security numbers, such as 111111111 and 123456789.
- Step 1.3.3:** Report the frequencies of likely invalid dates of birth, such as January 1, 1901 and November 11, 1911.

Stage 4: Standardize Last Name

- Step 1.4.1:** Remove last name suffixes that are contained within the last name field, rather than a distinct suffix field. *E.g.*, <Smith Jr.> becomes <Smith>.
- Step 1.4.2:** For last names containing hyphens, populate separate last name fields for all parts of the last name. *E.g.*, the last name <Smith-Jones> would have the value <Smith> entered into a LastName1 field and the value <Jones> entered into a LastName2 field.
- Step 1.4.3:** Remove spaces, hyphens, periods, and apostrophes from all last name fields and convert all letters to uppercase. *E.g.*, <O'Connor> becomes <OCONNOR> and <Smith-Jones> becomes <SMITHJONES>.
- Step 1.4.4:** Code all missing values as blank fields.

Exhibit A

Stage 5: Standardize First Name and Middle Name

Step 1.5.1: Remove spaces, hyphens, periods, and apostrophes from the first name field and convert all letters to uppercase. *E.g.*, <Jean-Paul> becomes <JEANPAUL>.

Step 1.5.2: Parse the first letter of the middle name (if available) and use it to populate a middle initial field. *E.g.*, <John> would yield <J>.¹

Step 1.5.3: Code all missing values as blank fields.

Stage 6: Standardize Date of Birth

Step 1.6.1: Convert the date of birth to an eight-digit string of MMDDYYYY.

Step 1.6.2: Code all missing values as blank fields.

Stage 7: Standardize Gender

Step 1.7.1: Code gender as a string of 1 for females and 0 for males.

Step 1.7.2: Fill missing gender values using the most common gender value for the first name associated with a record. *E.g.*, if 99% of records with first name <JOHN> are listed as male, assign the male identifier to all records with first name <JOHN> and no listed gender.

Step 1.7.3: If missing values remain, code all missing values as blank fields.

¹ The U.S. Department of State does not maintain a separate field for middle names in its database of U.S. Passport and Passport Card holders. Instead, both first and middle name may be stored in the first name field. For this database, the following rule will be applied: treat the first word in the first name field as the first name, and treat the first letter following the first space as the middle initial.

Exhibit A

Stage 8: Standardize Address

- Step 1.8.1:** Convert the residential ZIP code to a string if it is stored as a numeric field.
- Step 1.8.2:** Where the residential address ZIP code is blank, populate that field with the value in the mailing address ZIP code field, if available.²
- Step 1.8.3:** Truncate the residential ZIP code field to the first five digits. *E.g.*, <77777-1234> becomes <77777>.
- Step 1.8.4:** Where the residential address field is blank, populate that field with the value in the mailing address field, if available.
- Step 1.8.5:** Where address field containing street address begins with a street number, isolate the street number. *E.g.*, <123 Main Street> becomes <123>.
- Step 1.8.6:** Where the address field begins with recognized strings indicating a mail box, eliminate strings to isolate the box number. *E.g.*, <PO Box 444> becomes <444>.
- Step 1.8.7:** If missing values remain, code all missing values as blank fields.

Stage 9: Standardize Social Security Number

- Step 1.9.1:** Convert the social security number to a string if it is stored as a numeric field.
- Step 1.9.2:** Using full social security number, check for invalid SSNs. In the case of invalid SSNs, code as missing. *E.g.*, <123456789> becomes ◇.

² For purposes of this database matching protocol, the only address fields utilized with respect to data regarding U.S. Passports and U.S. Passport Cards are those regarding mailing addresses.

Exhibit A

Step 1.9.3: Extract the last four digits of full social security number as a four-character string and use them to populate a separate SSN4 field.

Step 1.9.4: Code all missing values as blank fields.

PART II: DATABASE PREPARATION

Stage 1: Construct Primary Identifier Variables for United States' One-to-Many Sweeps

- Step 2.1.1:** Create Combination A: First Name + Last Name + Gender + DOB + Residential ZIP + Residential Street Number. *E.g.*, the separate fields <JEAN>, <SMITH>, <0>, <01011950>, <77777>, and <123> are combined to a single field <JEANSMITH00101195077777123>.³
- Step 2.1.2:** Create Combination B: Last Name + Gender + DOB + Residential ZIP + Residential Street Number.
- Step 2.1.3:** Create Combination C: Gender + Date of Birth + Residential ZIP + Residential Street Number.
- Step 2.1.4:** Create Combination D: First Name + Last Name + Date of Birth + Residential ZIP + Residential Street Number.
- Step 2.1.5:** Create Combination E: First Name + Last Name + Gender + Residential ZIP + Residential Street Number.
- Step 2.1.6:** Create Combination F: First Name + Last Name + Gender + DOB.
- Step 2.1.7:** Create Combination M: Texas Driver License Number (where available).

³ For the U.S. Department of State only, the name portion of any combination is truncated if it is more than 32 characters long.

Exhibit A

Stage 2: Construct Secondary Identifier Variables for United States' One-to-Many Sweeps

- Step 2.2.1:** Create Combination G: First Name + Middle Initial + Last Name + Date of Birth.⁴
- Step 2.2.2:** Create Combination H: SSN4 + Date of Birth + Residential ZIP.
- Step 2.2.3:** Create Combination I: SSN4 + First Name + Last Name + Date of Birth.
- Step 2.2.4:** Create Combination K: First Name + Last Name 1 + Middle Initial + Date of Birth.
- Step 2.2.5:** Create Combination L: First Name + Last Name 2 + Middle Initial + Date of Birth.
- Step 2.2.6:** Full Social Security Number.

Stage 3: Construct Identifiers Used Only For Texas's Many-to-Many Sweeps

- Step 2.3.1:** Create Combination for Texas's Sweep 1: SSN4 + Last Name + DOB.
- Step 2.3.2:** Create Combination for Texas's Sweep 3: First Name + Last Name + DOB
- Step 2.3.3:** Create Combination for Texas's Sweep 4: First Name + Middle Initial + Last Name + DOB

Note: Combinations for Texas's Sweeps 1 and 3 do not already exist as pre-made fields in the TEAM database extract but instead must be created from the underlying TEAM database fields, in addition to being constructed on the Federal database side. Texas's Sweep 4 is equivalent to

⁴ Only for the State Department, create three further variations of Combination G created using the State Department's "LFMName" field which contains Last, First, and Middle Names, in that order, truncated to a maximum length of 32 characters. Combination G1 is DOB + LFMName; Combination G2 is DOB + First two words of LFMName; and Combination G3 is DOB + First two words of LFMName + First character of third word of LFMName.

Exhibit A

the combination for the United States' Combination G. Texas's Sweep 2 is on full 9 social security number.⁵

Stage 4: Establish Identifier Uniqueness For Combinations A - L

Step 2.4.1: Generate a field that establishes the uniqueness of each identifier variable.

For federal databases, for each combination A-L, generate a field that establishes uniqueness among only Texas records and a field that establishes uniqueness among nationwide records. *E.g.*, if only one record has the string <JEANSMITH01011950012377777> for Combination A, populate the uniqueness field for Combination A for that record as <1>. If four records have the string <JOHNSMITHA0101950> for Combination G, populate the uniqueness field for Combination G for each of those records as <2>, which indicates any number greater than one.

⁵ For purposes of matching to U.S. Department of State Passport and Passport Card holder data, Texas's Sweep 3 and Sweep 4 are as follows: Sweep 3: Last Name + First Name (restricted to 32 characters) + DOB. Sweep 4: Last Name + First Name + Middle Initial (restricted to 32 characters) + DOB.

Exhibit A

PART III: MATCH DATABASES

Stage 1: United States' Primary One-to-Many Matching Sweeps

- Step 3.1.1:** For each case in which Combination A is unique in the TEAM database, match Combination A against Combination A in the identifier or disability database. For federal databases, use only the subset of records with Texas addresses in the identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination A output field. Where there is a match, indicate the uniqueness of Combination A in the identifier or disability database in the Combination A output field (*e.g.*, in cases where there is one matching record in the Federal database, <1> should be inserted into the Combination A output field, while a <2> should be inserted into the Combination A output field if the TEAM record matched 2 or more records in the Federal database).
- Step 3.1.2:** Use the procedure in Step 3.1.1 to match Combination B, Combination C, Combination D, Combination E, and Combination F in the TEAM database against the equivalent combination field in the identifier or disability database.
- Step 3.1.3:** Use the procedure in Step 3.1.1 to match Combination M in the TEAM database against the equivalent combination field in the identifier databases produced by the State of Texas.

Exhibit A

Stage 2: United States' Secondary One-to-Many Matching Sweeps

Step 3.2.1: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), and where Combination G is unique in the TEAM database, match Combination G against Combination G in the identifier or disability database. For federal databases, use only the subset of records with Texas addresses in the identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination G output field. Where there is a match, indicate the uniqueness of Combination G in the identifier or disability database in the Combination G output field (*e.g.*, <1> if a unique match and <2> if matched to more than one record).

Step 3.2.2: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination H, Combination I, and complete social security number⁶ in the TEAM database against the equivalent combination/field in the identifier or disability database.

Step 3.2.3: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination K against Combination G, Combination K, and Combination L in the identifier or disability database.

⁶ The full social security number is not created as a separate "combination" as it is its own field stored within the TEAM database under the field name "ssn".

Exhibit A

Step 3.2.4: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination L against Combination G, Combination K, and Combination L in the identifier or disability database.

Stage 3: United States' Nationwide Federal Sweeps

Step 3.3.1: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, and where Combination F is unique, match Combination F against Combination F in the nationwide identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination F nationwide output field. Where there is a match, indicate the uniqueness of Combination F in the identifier or disability database in the Combination F nationwide output field (*e.g.*, <1> if a unique match and <2> if matched to more than one record).

Step 3.3.2: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination G, Combination I, and full social security number in the TEAM database against the equivalent combination/field in the nationwide identifier or disability database.

Step 3.3.3: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination

Exhibit A

K against Combination G, Combination K, and Combination L in the nationwide identifier or disability database.

Step 3.3.4: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination L against Combination G, Combination K, and Combination L in the nationwide identifier or disability database.⁷

Stage 4: Texas' Many-to-Many Nationwide Sweeps

Step 3.4.1 Regardless of whether the combination for Sweep 1 is unique in the TEAM database, match against the equivalent combination in a nationwide search of the Federal database.

⁷ Step 3.3.5 for the State Department only: Match the following Combination G variations from applicable State Department records, first to include only the subset of records with Texas addresses, and then to include all applicable U.S. Passport and Passport Card records nationwide (e.g., without Texas addresses), against the following fields from the TEAM database:

- Combination G1 to DOB + Last_fix + First_fix + Middle_name from the TEAM database;
- Combination G2 to DOB + Last_fix + First_fix from the TEAM database;
- Combination G2 to DOB + Last_fix + First word of First_name from the TEAM database;
- Combination G3 to DOB + Last_fix + First_fix + Middle_Initial from the TEAM database;
- Combination G3 to DOB + Last_fix + First word of First_name + Middle_Initial from the TEAM database; and
- Combination G3 to DOB + Last_fix + First word of First_name + First character of Second word of First_fix from the TEAM database.

Attempt matches for all TEAM records, regardless of whether they matched in any prior sweeps. Indicate <1> if a unique match and <2> if matched to more than one record.

Exhibit A

- Step 3.4.2** Regardless of whether full 9 social security number is unique in the TEAM database, for Sweep 2, match against the equivalent field in a nationwide search of the Federal database.
- Step 3.4.3** Regardless of whether the combination for Sweep 3 is unique in the TEAM database, match against the equivalent combination in a nationwide search of the Federal database.
- Step 3.4.4** Regardless of whether the combination for Sweep 4 (Combination G) is unique in the TEAM database, match against the equivalent combination in a nationwide search of the Federal database.

Note: For each of the Texas many-to-many sweeps:

- Indicate <1> if any TEAM combination matches a single combination in the Federal database
- Indicate <2> if any TEAM combination matches more than one record in the Federal database.
- Indicate <0> if no match is achieved.

Examples:

- If there are two TEAM records that have identical versions of the combination for Sweep 1, and there is one record in the Federal database that matches on this combination, both of the underlying TEAM records will have a matching output of <1> for Sweep 1.
- If there are three TEAM records that have identical versions of the combination for Sweep 3, and there are five records in the Federal database that match on that combination, the three TEAM records will each have a matching output of <2> for Sweep 3.

Exhibit B

Step 1.3.1 Diagnostics: Report the Frequency of Missing Values for Each Field

6, 781,544 Qualifying U.S. Passport/U.S. Passport Card records with 'TX' as the mailing address state

Field Name	Records with Missing Values
last_name	0
first_name	38
date_of_birth	0
gender_code	5
ssn	25,640
Last_fix	0
multi_lastname1	6,717,750
multi_lastname2	6,717,749
first_fix	38
middle_initial	929,454
fem_str	0
dob_str	0
zip5_num_res_alt	50
ssn4_str	25,640

92,776,332 Qualifying U.S. Passport/Passport Card records without 'TX' as the mailing address state

Field Name	Records with Missing Values
last_name	0
first_name	1,596
date_of_birth	2
gender_code	247
ssn	641,147
last_fix	0
multi_lastname1	91,615,428
multi_lastname2	91,615,406
first_fix	1,596
middle_initial	12,567,280
fem_str	16
dob_str	0
zip5_res_alt	757,630
str_num_res_alt	43,961

IN THE UNITED STATES DISTRICT COURT
FOR THE SOUTHERN DISTRICT OF TEXAS
CORPUS CHRISTI DIVISION

MARC VEASEY, *et al.*,

Plaintiffs,

v.

RICK PERRY, *et al.*,

Defendants.

Civil Action No. 2:13-cv-193 (NGR)

UNITED STATES OF AMERICA,

Plaintiff,

TEXAS LEAGUE OF YOUNG VOTERS
EDUCATION FUND, *et al.*,

Plaintiff-Intervenors,

TEXAS ASSOCIATION OF HISPANIC
COUNTY JUDGES AND COUNTY
COMMISSIONERS, *et al.*,

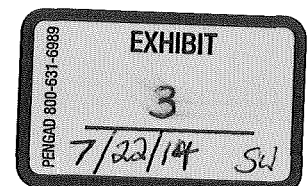
Plaintiff-Intervenors,

v.

STATE OF TEXAS, *et al.*,

Defendants.

Civil Action No. 2:13-cv-263 (NGR)



TEXAS STATE CONFERENCE OF NAACP
BRANCHES, *et al.*,

Plaintiffs,

v.

JOHN STEEN, *et al.*,

Defendants.

Civil Action No. 2:13-cv-291 (NGR)

BELINDA ORTIZ, *et al.*,

Plaintiffs,

v.

STATE OF TEXAS, *et al.*,

Defendants

Civil Action No. 2:13-cv-348 (NGR)

CORRECTED DECLARATION OF LEE CHARLES BAYDUSH

I, Lee Charles Baydush, pursuant to 28 U.S.C. § 1746, declare as follows:

1. I currently serve as subject matter expert for the U.S. Department of State, Bureau of Consular Affairs, Consular Systems & Technology, Service Integration & Innovation Division (CA/CST/SII). I have served Consular Affairs in this capacity with SII and others as a contractor specializing in Passport Systems for more than 28 years. My duties include research, analysis, consulting, and operational support of Consular IT systems.
2. I have personal knowledge of the information contained in this declaration based upon my work for the State Department, and my personal involvement in completion of the database comparison described below.

3. On February 27, 2014, the State Department received an encrypted hard drive containing data regarding Texas registered voters from the United States Department of Justice in the form of a 7 gigabyte .csv file containing 13,564,420 records for Texas registered voters. The State Department was able to load all records except for 6. The identifiers (VUIDs) for 5 of the 6 records that my agency was not able to load and analyze as part of the database comparison process are as follows: 1018504714; 1018510162; 1018555565; 1111894862; 1164253546. The final record that could not be loaded lacked a VUID, but has a unique identifier of 3890395vf (other than this unique identifier “var_fill” value, this record was blank).

4. The Texas data that the State Department received contained the following fields:

- Vuid	- dob_str
- last_name	- zip5_res_alt
- first_name	- str_num_res_alt
- middle_name	- AGDN
- ssn	- AGDN_TAG
- ssn4_str	- AGDN_unique
- gender_code	- AGDNlast
- official_id	- AGDNlast_TAG
- dl_id_str	- AGDNlast_unique
- date_of_birth	- AGD
- voter_status_code	- AGD_TAG
- block_number	- AGD_unique
- street_name	- and
- city	- ADN_TAG
- state	- ADN_unique
- zip_code	- AGN
- zip_str	- AGN_TAG
- mail_adrs_1	- AGN_unique
- last_fix	- GDN
- multi_lastname1	- GDN_TAG
- multi_lastname2	- GDN_unique
- first_fix	- DLN
- middle_initial	- DLN_TAG
- fem_str	- DLN_unique

- | | |
|----------------|-----------------|
| - unique_count | - FML1D_TAG |
| - FMLD | - FML1D_unique |
| - FMLD_TAG | - FML2D |
| - FMLD_unique | - FML2D_TAG |
| - SDZ | - FML2D_unique |
| - SDZ_TAG | - unique_count2 |
| - SDZ_unique | - var_fill |
| - SND | - K2 |
| - SND_TAG | - K3 |
| - SND_unique | - L2 |
| - FML1D | - L3 |

5. The State Department executed the data preparation and comparison steps, attached as Ex. A, that were provided by the United States Department of Justice.
6. Through this process, the State Department attempted to match particular combinations of identifying information for Texas registered voters (*e.g.*, first and last name, gender, and date of birth) with the same or related combinations of identifying information with respect to holders of U.S. Passports and U.S. Passport cards. For each particular combination, the State Department identified instances where the identifying information for a Texas registered voter matched with the respective combination of identifying information for one or more holders of a U.S. Passport or U.S. Passport Card.
7. For each sweep through the relevant State Department datasets, we appended a column to the Texas data to indicate, on a record-by-record basis, the output of the database comparisons on all of the requested combinations.
8. Two State Department datasets were created for the analysis—one contained qualifying records of U.S. Passport and U.S. Passport Card holders with a Texas mailing address and the other contained qualifying records of U.S. Passport and U.S. Passport Card holders without a Texas mailing address. The frequencies of any missing values in those datasets are provided as Ex. B.

9. On April 9, 2014, the State Department transferred responsive data to the Department of Justice, in the form of a variable width file with comma delimiters. In this file, there were 7,741 records in the TEAM database with commas embedded in the address field, and several hundred records with commas embedded in various name fields. Such commas could be incorrectly treated as delimiters in reading the Department of State matching results. As a result, I re-exported the results file using pipe delimiters and on May 14, 2014, the State Department transferred to the Department of Justice a compressed and encrypted 1.7 gigabyte zip file on a DVD that contained a 7.44 gigabyte pipe-delimited file. Both the April 9 and May 14 files are identical other than the use of different delimiters. The use of pipe delimiters on the May 14, 2014, file will prevent any difficulties reading the embedded commas in the results file. The name of the May 14, 2014 results file is Team_Pipe_Delim.
10. The results of each sweep, as described in Ex. A, can be found in the following columns of the data that the State Department returned to the Department of Justice on May 14:
 - a. "AGDN_matches_id" contains results of the Stage 1, Step 3.1.1, Combination A match;
 - b. "AGDNlast_matches_id" contains results of the Stage 1, Step 3.1.2, Combination B match;
 - c. "AGD_matches_id" contains results of the Stage 1, Step 3.1.2, Combination C match;
 - d. "ADN_matches_id" contains results of the Stage 1, Step 3.1.2, Combination D match;
 - e. "AGN_matches_id" contains results of the Stage 1, Step 3.1.2, Combination E match;
 - f. "GDN_matches_id" contains results of the Stage 1, Step 3.1.2, Combination F match;

- g. "FMLD_matches_id" contains results of the Stage 2, Step 3.2.1, Combination G match;
- h. "SDZ_matches_id" contains results of the Stage 2, Step 3.2.2, Combination H match;
- i. "SND_matches_id" contains results of the Stage 2, Step 3.2.2, Combination I match;
- j. "ssn_matches_id" contains results of the Stage 2, Step 3.2.2, nine-digit social security number match;
- k. "k2_matches_id" contains results of the Stage 2, Step 3.2.3, Combination K to G match;
- l. "FML1D_matches_id" contains results of the Stage 2, Step 3.2.3, Combination K match;
- m. "k3_matches_id" contains results of the Stage 2, Step 3.2.3, Combination K to L match;
- n. "l2_matches_id" contains results of the Stage 2, Step 3.2.4, Combination L to G match;
- o. "FML2D_matches_id" contains results of the Stage 2, Step 3.2.4, Combination L match;
- p. "l3_matches_id" contains results of the Stage 2, Step 3.2.4, Combination L to K match;
- q. "GDN_matches_id_Other_GDN" contains results of the Stage 3, Step 3.3.1, Combination F match;
- r. "FMLD_matches_id_Other_FMLD" contains results of the Stage 3, Step 3.3.2, Combination G match;

- s. "SND_matches_id_Other_SND" contains results of the Stage 3, Step 3.3.2, Combination I match;
- t. "ssn_matches_id_Other" contains results of the Stage 3, Step 3.3.2, nine-digit social security number match;
- u. "FML1D_matches_id_Other_FMLD" contains results of the Stage 3, Step 3.3.3, Combination K to G match;
- v. "FML1D_matches_id_Other_FML1D" contains results of the Stage 2, Step 3.3.3, Combination K match;
- w. "FML1D_matches_id_Other_FML2D" contains results of the Stage 2, Step 3.3.3, Combination K to L match;
- x. "FML2D_matches_id_Other_FMLD" contains results of the Stage 3, Step 3.3.4, Combination L to G match;
- y. "FML2D_matches_id_Other_FML2D" contains results of the Stage 3, Step 3.3.4, Combination L match;
- z. "FML2D_matches_id_Other_FML1D" contains results of the Stage 3, Step 3.3.4, Combination L to K match;
- aa. "a1_matches_id_TX" and "a1_matches_id_Other" respectively contain the results of the Stage 3, Step 3.3.5 Combination G1 to DOB + Last_fix + First_fix + Middle_name matches for the Texas-only and non-Texas address datasets;
- bb. "a2_1_matches_id_TX" and "a2_1_matches_id_Other" respectively contain the results of the Stage 3, Step 3.3.5 Combination G2 to DOB + Last_fix + First_fix matches for the Texas-only and non-Texas address datasets;

- cc. "a2_2_matches_id_TX" and "a2_2_matches_id_Other" respectively contain the results of the Stage 3, Step 3.3.5 Combination G2 to DOB + Last_fix + First word of First_name matches for the Texas-only and non-Texas address datasets;
- dd. "a3_1_matches_id_TX" and "a3_1_matches_id_Other" respectively contain the results of the Stage 3, Step 3.3.5 Combination G3 to DOB + Last_fix + First_fix + Middle_Initial matches for the Texas-only and non-Texas address datasets;
- ee. "a3_2_matches_id_TX" and "a3_2_matches_id_Other" respectively contain the results of the Stage 3, Step 3.3.5 Combination G3 to DOB Last_fix + First word of First_name + Middle_Initial matches for the Texas-only and non-Texas address datasets;
- ff. "a3_3_matches_id_TX" and "a3_3_matches_id_Other" respectively contain the results of the Stage 3, Step 3.3.5 Combination G3 to DOB Last_fix + First word of First_name + First character of Second word of First_fix matches for the Texas-only and non-Texas address datasets;
- gg. "TXMM_snlastdAll_matches_id" contains results of the Stage 4, Step 3.4.1, Sweep 1 match, while "TXMM_TX_snlastdAll_matches_id" contains the results for the Sweep 1 match based on the Texas-only dataset and "TXMM_Other_snlastdAll_matches_id" contains the results for the Sweep 1 match based on the non-Texas addresses dataset;
- hh. "TXMM_ssn_matches_id" contains results of the Stage 4, Step 3.4.2, Sweep 2 match, while "TXMM_TX_ssn_matches_id" contains the results for the Sweep 2 match based on the Texas-only dataset and "TXMM_Other_ssn_matches_id" contains the results for the Sweep 2 match based on the non-Texas address dataset;

- ii. "TXMM_ndAll_matches_id" contains results of the Stage 4, Step 3.4.3, Sweep 3 match, while "TXMM_TX_ndAll_matches_id" contains the results for the Sweep 3 match based on the Texas-only dataset and "TXMM_Other_ndAll_matches_id" contains the results for the Sweep 3 match based on the non-Texas address dataset; and
- jj. "TXMM_fmldAll_matches_id" contains results of the Stage 4, Step 3.4.4, Sweep 4 match, while "TXMM_TX_fmldAll_matches_id" contains the results for the Sweep 4 match based on the Texas-only dataset and "TXMM_Other_fmldAll_matches_id" contains the results for the Sweep 4 match based on the non-Texas address dataset.

11. No other matching comparisons between the Texas data and data from State Department were undertaken beyond those set forth above, in accordance with the steps provided in Ex. A.

I declare under penalty of perjury that the foregoing is true and correct.

Executed on June 24, 2014.

Lu Baydush

Exhibit A

United States v. Texas: Federal Agency Algorithm Summary

This document summarizes the database matches that both the United States and all other plaintiffs and the State of Texas have requested from Federal Agencies as part of the *Veasey v. Perry/United States v. Texas* litigation (S.D. Tex).

The matching process proceeds in three parts, which are explained in detail below. *First*, databases are prepared and standardized. *Second*, identifier values are constructed by combining multiple individual fields. *Third*, the United States' one-to-many matches and the State of Texas's many-to-many matches are conducted between databases.

PART I: DATABASE PREPARATION

Stage 1: Extraction of Available Data from Federal Identification & Disability Databases

- Step 1.1.1:** Extract complete name into separate first name, middle name, and last name fields.
- Step 1.1.2:** Extract date of birth.
- Step 1.1.3:** Extract gender.
- Step 1.1.4:** Extract residential address and mailing address.
- Step 1.1.5:** Extract social security number.
- Step 1.1.6:** Extract Texas driver license number (only if present in Federal database).

Stage 2: Separate Valid Identification and Disability Records

- Step 1.2.1:** Remove records from identification database extracts that indicate that an ID has been revoked or has expired more than 60 days before the date of the TEAM database snapshot (which is January 15, 2014).

Exhibit A

Step 1.2.2: Remove records from disability database extracts that do not indicate current disability status or indicate a Veterans Administration disability rating of less than 50%.

Stage 3: Diagnostics

Step 1.3.1: Report the frequency of missing values for each field.

Step 1.3.2: Report the frequencies of invalid Social Security numbers, such as 111111111 and 123456789.

Step 1.3.3: Report the frequencies of likely invalid dates of birth, such as January 1, 1901 and November 11, 1911.

Stage 4: Standardize Last Name

Step 1.4.1: Remove last name suffixes that are contained within the last name field, rather than a distinct suffix field. *E.g.*, <Smith Jr.> becomes <Smith>.

Step 1.4.2: For last names containing hyphens, populate separate last name fields for all parts of the last name. *E.g.*, the last name <Smith-Jones> would have the value <Smith> entered into a LastName1 field and the value <Jones> entered into a LastName2 field.

Step 1.4.3: Remove spaces, hyphens, periods, and apostrophes from all last name fields and convert all letters to uppercase. *E.g.*, <O'Connor> becomes <OCONNOR> and <Smith-Jones> becomes <SMITHJONES>.

Step 1.4.4: Code all missing values as blank fields.

Exhibit A

Stage 5: Standardize First Name and Middle Name

- Step 1.5.1:** Remove spaces, hyphens, periods, and apostrophes from the first name field and convert all letters to uppercase. *E.g.*, <Jean-Paul> becomes <JEANPAUL>.
- Step 1.5.2:** Parse the first letter of the middle name (if available) and use it to populate a middle initial field. *E.g.*, <John> would yield <J>.¹
- Step 1.5.3:** Code all missing values as blank fields.

Stage 6: Standardize Date of Birth

- Step 1.6.1:** Convert the date of birth to an eight-digit string of MMDDYYYY.
- Step 1.6.2:** Code all missing values as blank fields.

Stage 7: Standardize Gender

- Step 1.7.1:** Code gender as a string of 1 for females and 0 for males.
- Step 1.7.2:** Fill missing gender values using the most common gender value for the first name associated with a record. *E.g.*, if 99% of records with first name <JOHN> are listed as male, assign the male identifier to all records with first name <JOHN> and no listed gender.
- Step 1.7.3:** If missing values remain, code all missing values as blank fields.

¹ The U.S. Department of State does not maintain a separate field for middle names in its database of U.S. Passport and Passport Card holders. Instead, both first and middle name may be stored in the first name field. For this database, the following rule will be applied: treat the first word in the first name field as the first name, and treat the first letter following the first space as the middle initial.

Exhibit A

Stage 8: Standardize Address

- Step 1.8.1:** Convert the residential ZIP code to a string if it is stored as a numeric field.
- Step 1.8.2:** Where the residential address ZIP code is blank, populate that field with the value in the mailing address ZIP code field, if available.²
- Step 1.8.3:** Truncate the residential ZIP code field to the first five digits. *E.g.*, <77777-1234> becomes <77777>.
- Step 1.8.4:** Where the residential address field is blank, populate that field with the value in the mailing address field, if available.
- Step 1.8.5:** Where address field containing street address begins with a street number, isolate the street number. *E.g.*, <123 Main Street> becomes <123>.
- Step 1.8.6:** Where the address field begins with recognized strings indicating a mail box, eliminate strings to isolate the box number. *E.g.*, <PO Box 444> becomes <444>.
- Step 1.8.7:** If missing values remain, code all missing values as blank fields.

Stage 9: Standardize Social Security Number

- Step 1.9.1:** Convert the social security number to a string if it is stored as a numeric field.
- Step 1.9.2:** Using full social security number, check for invalid SSNs. In the case of invalid SSNs, code as missing. *E.g.*, <123456789> becomes <>.

² For purposes of this database matching protocol, the only address fields utilized with respect to data regarding U.S. Passports and U.S. Passport Cards are those regarding mailing addresses.

Exhibit A

Step 1.9.3: Extract the last four digits of full social security number as a four-character string and use them to populate a separate SSN4 field.

Step 1.9.4: Code all missing values as blank fields.

PART II: DATABASE PREPARATION

Stage 1: Construct Primary Identifier Variables for United States' One-to-Many Sweeps

- Step 2.1.1:** Create Combination A: First Name + Last Name + Gender + DOB + Residential ZIP + Residential Street Number. *E.g.*, the separate fields <JEAN>, <SMITH>, <0>, <01011950>, <77777>, and <123> are combined to a single field <JEANSMITH00101195077777123>.³
- Step 2.1.2:** Create Combination B: Last Name + Gender + DOB + Residential ZIP + Residential Street Number.
- Step 2.1.3:** Create Combination C: Gender + Date of Birth + Residential ZIP + Residential Street Number.
- Step 2.1.4:** Create Combination D: First Name + Last Name + Date of Birth + Residential ZIP + Residential Street Number.
- Step 2.1.5:** Create Combination E: First Name + Last Name + Gender + Residential ZIP + Residential Street Number.
- Step 2.1.6:** Create Combination F: First Name + Last Name + Gender + DOB.
- Step 2.1.7:** Create Combination M: Texas Driver License Number (where available).

³ For the U.S. Department of State only, the name portion of any combination is truncated if it is more than 32 characters long.

Exhibit A

Stage 2: Construct Secondary Identifier Variables for United States' One-to-Many Sweeps

- Step 2.2.1:** Create Combination G: First Name + Middle Initial + Last Name + Date of Birth.⁴
- Step 2.2.2:** Create Combination H: SSN4 + Date of Birth + Residential ZIP.
- Step 2.2.3:** Create Combination I: SSN4 + First Name + Last Name + Date of Birth.
- Step 2.2.4:** Create Combination K: First Name + Last Name 1 + Middle Initial + Date of Birth.
- Step 2.2.5:** Create Combination L: First Name + Last Name 2 + Middle Initial + Date of Birth.
- Step 2.2.6:** Full Social Security Number.

Stage 3: Construct Identifiers Used Only For Texas's Many-to-Many Sweeps

- Step 2.3.1:** Create Combination for Texas's Sweep 1: SSN4 + Last Name + DOB.
- Step 2.3.2:** Create Combination for Texas's Sweep 3: First Name + Last Name + DOB
- Step 2.3.3:** Create Combination for Texas's Sweep 4: First Name + Middle Initial + Last Name + DOB

Note: Combinations for Texas's Sweeps 1 and 3 do not already exist as pre-made fields in the TEAM database extract but instead must be created from the underlying TEAM database fields, in addition to being constructed on the Federal database side. Texas's Sweep 4 is equivalent to

⁴ Only for the State Department, create three further variations of Combination G created using the State Department's "LFMName" field which contains Last, First, and Middle Names, in that order, truncated to a maximum length of 32 characters. Combination G1 is DOB + LFMName; Combination G2 is DOB + First two words of LFMName; and Combination G3 is DOB + First two words of LFMName + First character of third word of LFMName.

Exhibit A

the combination for the United States' Combination G. Texas's Sweep 2 is on full 9 social security number.⁵

Stage 4: Establish Identifier Uniqueness For Combinations A - L

Step 2.4.1: Generate a field that establishes the uniqueness of each identifier variable.

For federal databases, for each combination A-L, generate a field that establishes uniqueness among only Texas records and a field that establishes uniqueness among only Texas records and a field that establishes uniqueness among nationwide records. *E.g.*, if only one record has the string <JEANSMITH01011950012377777> for Combination A, populate the uniqueness field for Combination A for that record as <1>. If four records have the string <JOHNSMITHA0101950> for Combination G, populate the uniqueness field for Combination G for each of those records as <2>, which indicates any number greater than one.

⁵ For purposes of matching to U.S. Department of State Passport and Passport Card holder data, Texas's Sweep 3 and Sweep 4 are as follows: Sweep 3: Last Name + First Name (restricted to 32 characters) + DOB. Sweep 4: Last Name + First Name + Middle Initial (restricted to 32 characters) + DOB.

Exhibit A

PART III: MATCH DATABASES

Stage 1: United States' Primary One-to-Many Matching Sweeps

- Step 3.1.1:** For each case in which Combination A is unique in the TEAM database, match Combination A against Combination A in the identifier or disability database. For federal databases, use only the subset of records with Texas addresses in the identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination A output field. Where there is a match, indicate the uniqueness of Combination A in the identifier or disability database in the Combination A output field (*e.g.*, in cases where there is one matching record in the Federal database, <1> should be inserted into the Combination A output field, while a <2> should be inserted into the Combination A output field if the TEAM record matched 2 or more records in the Federal database).
- Step 3.1.2:** Use the procedure in Step 3.1.1 to match Combination B, Combination C, Combination D, Combination E, and Combination F in the TEAM database against the equivalent combination field in the identifier or disability database.
- Step 3.1.3:** Use the procedure in Step 3.1.1 to match Combination M in the TEAM database against the equivalent combination field in the identifier databases produced by the State of Texas.

Exhibit A

Stage 2: United States' Secondary One-to-Many Matching Sweeps

Step 3.2.1: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), and where Combination G is unique in the TEAM database, match Combination G against Combination G in the identifier or disability database. For federal databases, use only the subset of records with Texas addresses in the identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination G output field. Where there is a match, indicate the uniqueness of Combination G in the identifier or disability database in the Combination G output field (*e.g.*, <1> if a unique match and <2> if matched to more than one record).

Step 3.2.2: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination H, Combination I, and complete social security number⁶ in the TEAM database against the equivalent combination/field in the identifier or disability database.

Step 3.2.3: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination K against Combination G, Combination K, and Combination L in the identifier or disability database.

⁶ The full social security number is not created as a separate “combination” as it is its own field stored within the TEAM database under the field name “ssn”.

Exhibit A

Step 3.2.4: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination L against Combination G, Combination K, and Combination L in the identifier or disability database.

Stage 3: United States' Nationwide Federal Sweeps

Step 3.3.1: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, and where Combination F is unique, match Combination F against Combination F in the nationwide identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination F nationwide output field. Where there is a match, indicate the uniqueness of Combination F in the identifier or disability database in the Combination F nationwide output field (*e.g.*, <1> if a unique match and <2> if matched to more than one record).

Step 3.3.2: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination G, Combination I, and full social security number in the TEAM database against the equivalent combination/field in the nationwide identifier or disability database.

Step 3.3.3: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination

Exhibit A

K against Combination G, Combination K, and Combination L in the nationwide identifier or disability database.

Step 3.3.4: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination L against Combination G, Combination K, and Combination L in the nationwide identifier or disability database.⁷

Stage 4: Texas' Many-to-Many Nationwide Sweeps

Step 3.4.1 Regardless of whether the combination for Sweep 1 is unique in the TEAM database, match against the equivalent combination in a nationwide search of the Federal database.

⁷ Step 3.3.5 for the State Department only: Match the following Combination G variations from applicable State Department records, first to include only the subset of records with Texas addresses, and then to include all applicable U.S. Passport and Passport Card records nationwide (*e.g.*, without Texas addresses), against the following fields from the TEAM database:

- Combination G1 to DOB + Last_fix + First_fix + Middle_name from the TEAM database;
- Combination G2 to DOB + Last_fix + First_fix from the TEAM database;
- Combination G2 to DOB + Last_fix + First word of First_name from the TEAM database;
- Combination G3 to DOB + Last_fix + First_fix + Middle_Initial from the TEAM database;
- Combination G3 to DOB + Last_fix + First word of First_name + Middle_Initial from the TEAM database; and
- Combination G3 to DOB + Last_fix + First word of First_name + First character of Second word of First_fix from the TEAM database.

Attempt matches for all TEAM records, regardless of whether they matched in any prior sweeps. Indicate <1> if a unique match and <2> if matched to more than one record.

Exhibit A

- Step 3.4.2** Regardless of whether full 9 social security number is unique in the TEAM database, for Sweep 2, match against the equivalent field in a nationwide search of the Federal database.
- Step 3.4.3** Regardless of whether the combination for Sweep 3 is unique in the TEAM database, match against the equivalent combination in a nationwide search of the Federal database.
- Step 3.4.4** Regardless of whether the combination for Sweep 4 (Combination G) is unique in the TEAM database, match against the equivalent combination in a nationwide search of the Federal database.

Note: For each of the Texas many-to-many sweeps:

- Indicate <1> if any TEAM combination matches a single combination in the Federal database
- Indicate <2> if any TEAM combination matches more than one record in the Federal database.
- Indicate <0> if no match is achieved.

Examples:

- If there are two TEAM records that have identical versions of the combination for Sweep 1, and there is one record in the Federal database that matches on this combination, both of the underlying TEAM records will have a matching output of <1> for Sweep 1.
- If there are three TEAM records that have identical versions of the combination for Sweep 3, and there are five records in the Federal database that match on that combination, the three TEAM records will each have a matching output of <2> for Sweep 3.

Exhibit B

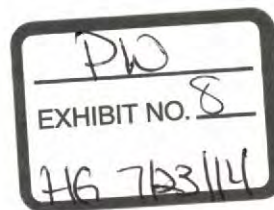
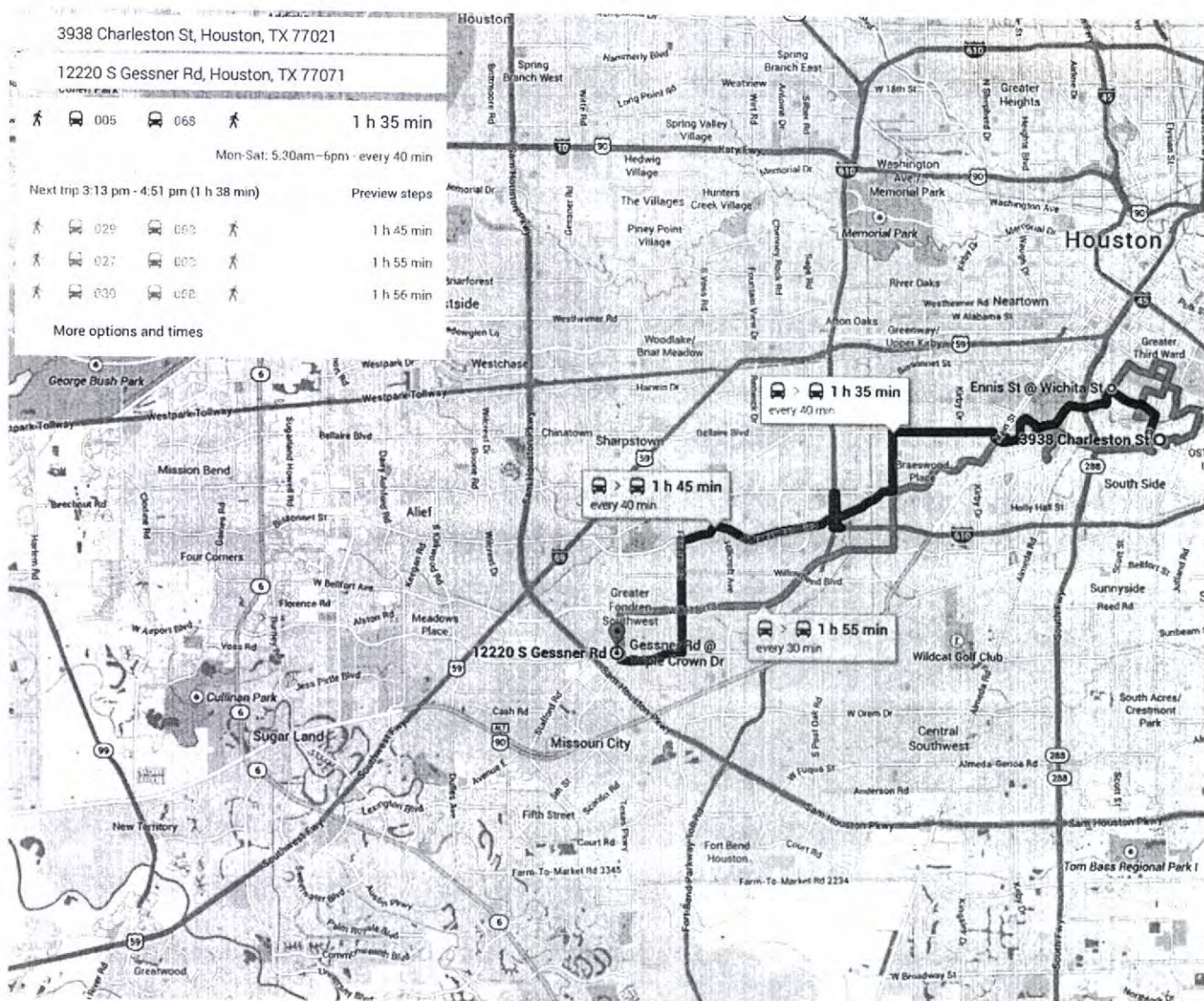
Step 1.3.1 Diagnostics: Report the Frequency of Missing Values for Each Field

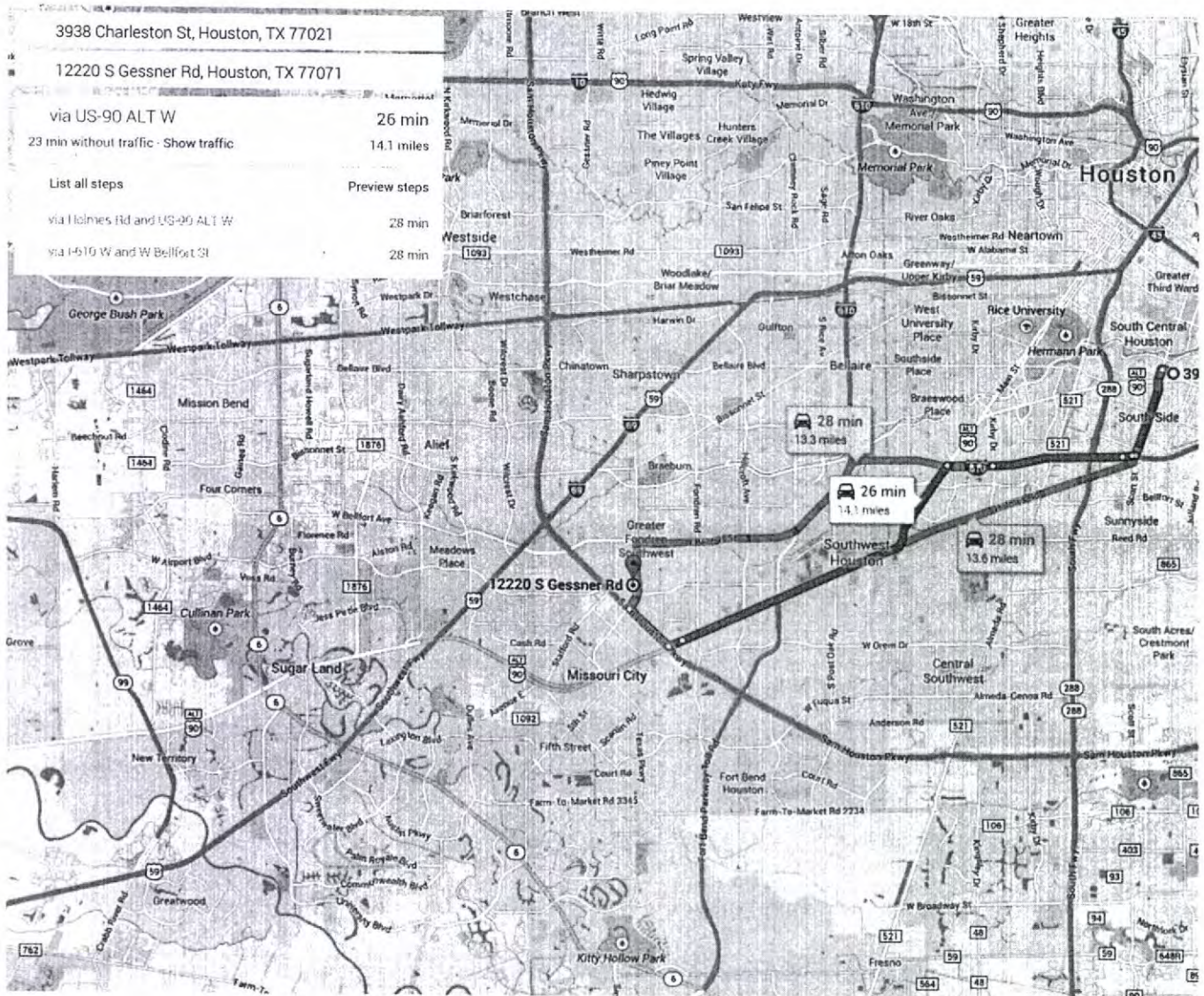
6, 781,544 Qualifying U.S. Passport/U.S. Passport Card records with 'TX' as the mailing address state

Field Name	Records with Missing Values
last_name	0
first_name	38
date_of_birth	0
gender_code	5
ssn	25,640
Last_fix	0
multi_lastname1	6,717,750
multi_lastname2	6,717,749
first_fix	38
middle_initial	929,454
fem_str	0
dob_str	0
zip5_num_res_alt	50
ssn4_str	25,640

92,776,332 Qualifying U.S. Passport/Passport Card records without 'TX' as the mailing address state

Field Name	Records with Missing Values
last_name	0
first_name	1,596
date_of_birth	2
gender_code	247
ssn	641,147
last_fix	0
multi_lastname1	91,615,428
multi_lastname2	91,615,406
first_fix	1,596
middle_initial	12,567,280
fem_str	16
dob_str	0
zip5_res_alt	757,630
str_num_res_alt	43,961





NDLS

Page 1 of 2

Selected Applicant: WASHINGTON, PHYLLIS DL/ID/UNL: 25192834 DOB: 07/23/1947

PL917
9/2/2014
2:13-cv-00193

WASHINGTON, PHYLLIS
MARIE

Transaction Eligibility
ID DL
Other



DL/ID/UNL Number: 25192834
Date of Birth: 07/23/1947
Sex: FEMALE
Eye Color: BROWN
Record Status: ELIGIBLE

PERSONAL	ADDRESSES	CARDS	EDUCATION	AKA	ID DOCUMENTS	MICROFILM
DOCUMENTS	OTHER					

- Basic Information

First Name: PHYLLIS
Middle Name: MARIE
Last Name: WASHINGTON
Suffix:
SSN: 449-86-5849
SSN Affidavit: NO
DOB: 07/23/1947
Age: 66
UNL Number:

- Primary Emergency Contact

First Name:
Last Name:
Phone 1:
Phone 2:
Street 1:
City:
State:
Zip Code:
Country:

- Physical Description

Sex: FEMALE
Eye Color: BROWN
Hair Color: BLACK
Race: BLACK
Height: 5' 04"
Weight: 190

- Secondary Emergency Contact

First Name:
Last Name:
Phone 1:
Phone 2:
Street 1:
City:
State:
Zip Code:
Country:

- Place of Birth Information

Country of Birth:
State of Birth:
City of Birth:
County of Birth:

- Other Information

Record Creation Date: 11/14/2006
Communication Impediment: NO
Organ Donor: NO
Disabled Veteran: NO
Veteran: NO
US Citizen: YES
Military: NO



NDLS

Page 2 of 2

-Home Bound:	NO
Out Of State:	NO
Valid Without Photo:	NO
Voter Registration:	NO
Selective Service:	NO
Limited Term:	NO
Limited Term Expiration Date:	
Combined Date:	
CDLIS Indicator:	NO
PDPS Indicator:	NO

NDLS

Page 1 of 2

Selected Applicant: WASHINGTON, PHYLLIS DL/ID/UNL: 25192834 DOB: 07/23/1947

**WASHINGTON, PHYLLIS
MARIE****Transaction Eligibility****ID
Other****DL****DL/ID/UNL Number:** 25192834**Date of Birth:** 07/23/1947**Sex:** FEMALE**Eye Color:** BROWN**Record Status:** ELIGIBLE

PERSONAL	ADDRESSES	CARDS	EDUCATION	AKA	ID DOCUMENTS	MICROFILM
DOCUMENTS	OTHER					

- Physical Address

Street 1: 3938 CHARLESTON ST
Street 2:
City: HOUSTON
State: TEXAS
Zip Code: 77021
Zip Code Extension: 0000
County: HARRIS
Country: UNITED STATES

- Mailing Address

Street 1: 3938 CHARLESTON ST
Street 2:
City: HOUSTON
State: TEXAS
Zip Code: 77021
Zip Code Extension: 0000
Country: UNITED STATES

- Special Mailing Address

In Care Of:
Street 1:
Street 2:
City:
State:
Zip Code:
Zip Code Extension:
Country:

- Applicant Contact Information

NDLS

Page 2 of 2

Home Phone:

~~(832) 245-4232~~

Other Phone:

Email:

NDLS

Page 1 of 1

Selected Applicant: WASHINGTON, PHYLLIS DL/ID/UNL: 25192834 DOB: 07/23/1947

**WASHINGTON, PHYLLIS
MARIE**

Transaction Eligibility

ID **DL**
Other



DL/ID/UNL Number: 25192834
Date of Birth: 07/23/1947
Sex: FEMALE
Eye Color: BROWN
Record Status: ELIGIBLE

PERSONAL	ADDRESSES	CARDS	EDUCATION	AKA	ID DOCUMENTS	MICROFILM
DOCUMENTS	OTHER					

ID Card Info

ID Number:	25192834
Original Issue Date:	11/14/2006
Issuance Status:	ACTIVE - OUTSIDE RENEWAL PERIOD
Issue Date:	12/03/2013
SO Issuance ID:	None
Audit Number:	21210371121053247568
Expiration Date:	INDEFINITE

NDLS

Page 1 of 1

Selected Applicant: WASHINGTON, PHYLLIS DL/ID/UNL: 25192834 DOB: 07/23/1947

**WASHINGTON, PHYLLIS
MARIE****Transaction Eligibility**

ID	DL
Other	

**DL/ID/UNL Number:** 25192834**Date of Birth:** 07/23/1947**Sex:** FEMALE**Eye Color:** BROWN**Record Status:** ELIGIBLE

PERSONAL	ADDRESSES	CARDS	EDUCATION	AKA	ID DOCUMENTS	MICROFILM
DOCUMENTS	OTHER					

- Proof Of ID Document Information

Document Type	Document Name	Document Number	Expiration Date
PRIMARY	TEXAS DL/ID (WITHIN 2 YEARS OF EXPIRATION)	25192834	

- SSN Document Information

Document Name	Document Number
SOCIAL SECURITY CARD	449865849

- Surrendered Document Information

No History of Surrendered Cards on Record.

NDLS

Page 1 of 1

Selected Applicant: WASHINGTON, PHYLLIS DL/ID/UNL: 25192834 DOB: 07/23/1947

**WASHINGTON, PHYLLIS
MARIE**

Transaction Eligibility

ID ☐ DL ☐
Other ☐



DL/ID/UNL Number: 25192834

Date of Birth: 07/23/1947

Sex: FEMALE

Eye Color: BROWN

Record Status: ELIGIBLE

PERSONAL	ADDRESSES	CARDS	EDUCATION	AKA	ID DOCUMENTS	MICROFILM
DOCUMENTS	OTHER					

- Other Information:

Texas FTA/FTP: NO

OOS FTA/FTP: NO

Total DRP Points: 0

- Sex Offender Information

Sex Offender: NO

SO Registration Date:

SID Number:

CSO Date Of Notice:

CSO Public:

- SSOLV Information

Social Security Number: 449-86-5849

SSN Affidavit: NO

SSOLV Result: SSN VERIFIED

SSOLV Transaction Status: VERIFIED

SSN Audit:

NDLS

Page 1 of 1

Selected Applicant: WASHINGTON, PHYLLIS DL/ID/UNL: 25192834 DOB: 07/23/1947

**WASHINGTON, PHYLLIS
MARIE****Transaction Eligibility**ID
Other

DL

**DL/ID/UNL Number:** 25192834**Date of Birth:** 07/23/1947**Sex:** FEMALE**Eye Color:** BROWN**Record Status:** ELIGIBLE

PERSONAL	ADDRESSES	CARDS	EDUCATION	AKA	ID DOCUMENTS	MICROFILM
DOCUMENTS	OTHER					

- Other Information:

Texas FTA/FTP: NO

OOS FTA/FTP: NO

Total DRP Points: 0

- Sex Offender Information

Sex Offender: NO

SO Registration Date:

SID Number:

CSO Date Of Notice:

CSO Public:

- SSOLV Information

Social Security Number: 449-86-5849

SSN Affidavit: NO

SSOLV Result: SSN VERIFIED

SSOLV Transaction Status: VERIFIED

SSN Audit:

NDLS

Page 1 of 1

Selected Applicant: WASHINGTON, PHYLLIS DLID/UNL: 25192834 DOB: 07/23/1947

Transaction History

Administrative Update Date:

Transaction Type	Transaction Status	DLID/UNL Number	Application Date	Issue Date	Expiration Date	Trace Number	CNF (Correction - No Fee)	Request Method	Location	User ID	Details
Renew ID	Mail Complete	25192834	12/03/2013 11:43:33	12/03/2013	INDEFINITE	405DLA020896158	NO	FIELD	HOUSTON - GESSNER (SUPER)	PW15333	Production Details Lawful Presence
Duplicate ID	Mail Complete	25192834	01/09/2008 16:20:33	01/09/2008		00072950008	NO	FIELD	HOUSTON - GESSNER (SUPER)	E210	Production Details
Original ID	Mail Complete	25192834	11/15/2006 16:20:33	11/15/2006		20063182170302840001	NO	FIELD	HOUSTON - TOWNHURST	MIGRATION	Production Details Lawful Presence

IN THE UNITED STATES DISTRICT COURT
FOR THE SOUTHERN DISTRICT OF TEXAS
CORPUS CHRISTI DIVISION

MARC VEASEY, *et al.*,

Plaintiffs,

v.

RICK PERRY, *et al.*,

Defendants.

Civil Action No. 2:13-cv-193 (NGR)

UNITED STATES OF AMERICA,

Plaintiff,

TEXAS LEAGUE OF YOUNG VOTERS
EDUCATION FUND, *et al.*,

Plaintiff-Intervenors,

TEXAS ASSOCIATION OF HISPANIC
COUNTY JUDGES AND COUNTY
COMMISSIONERS, *et al.*,

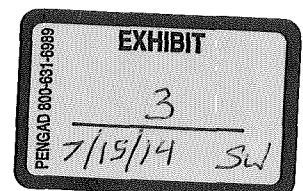
Plaintiff-Intervenors,

v.

STATE OF TEXAS, *et al.*,

Defendants.

Civil Action No. 2:13-cv-263 (NGR)



TEXAS STATE CONFERENCE OF NAACP
BRANCHES, *et al.*,

Plaintiffs,

v.

JOHN STEEN, *et al.*,

Defendants.

Civil Action No. 2:13-cv-291 (NGR)

BELINDA ORTIZ, *et al.*,

Plaintiffs,

v.

STATE OF TEXAS, *et al.*,

Defendants

Civil Action No. 2:13-cv-348 (NGR)

DECLARATION OF STEVE STRAUSLER

I, Steve Strausler, pursuant to 28 U.S.C. § 1746, declare as follows:

1. I currently serve as Computer Specialist with the Social Security Administration ("SSA"). I have served in this capacity for 33 years. My duties include updating and maintenance of SSA's Master Beneficiary Record, System of Records Number 60-0090 (published at 71 F.R. 1826 (January 11, 2006)).
2. I have personal knowledge of the information contained in this declaration based upon my work for SSA, and my personal involvement in completion of the database comparison described below.

3. On February 28, 2014, SSA received from the United States Department of Justice (“DOJ”) an encrypted hard drive containing data regarding Texas registered voters in the form of a 7 gigabyte .csv file containing 13,564,420 records for Texas registered voters. SSA was unable to use this variable width file. We requested that the DOJ provide a fixed width file, which we received from DOJ on April 3, 2014. 13,564,419 records were provided in the fixed width file and SSA was able to load all of them.¹ In order to hold down the file size and make the fixed width file more manageable for SSA to load, this file contained only the fields that were strictly necessary to conduct the database comparison process, rather than all fields that were used to construct the matching combinations, as had been sent in the initial variable width file.
4. The Texas data on the 9.4 GB fixed width file that SSA received contained the following fields:

¹ One record from the variable width file was omitted because of data formatting issues with that record.

- | | |
|------------------|------------|
| - vuid | - GDN |
| - ssn | - DLN |
| - ssn4_str | - FMLD |
| - zip_str | - SDZ |
| - last_fix | - SND |
| - first_fix | - FML1D |
| - middle_initial | - FML2D |
| - dob_str | - var_fill |
| - AGDN | - K2 |
| - AGDNlast | - K3 |
| - AGD | - L2 |
| - ADN | - L3 |
| - AGN | |

5. SSA executed the data preparation and comparison steps, attached as Ex. A, that were provided by the DOJ.
6. Through this process, SSA attempted to match particular combinations of identifying information for Texas registered voters (*e.g.*, first and last name, gender, and date of birth) with the same or related combinations of identifying information with respect to records from SSA's master beneficiary records² of persons who are 18 years of age or older, and who SSA has determined to be in current payment status for disability.³
7. For each sweep through the relevant SSA data, we appended results to the Texas data to indicate, on a record-by-record basis, the output of the database comparisons on all of the requested combinations. The frequencies of missing values in the underlying SSA data are provided as Ex. B.

² SSA's Master Beneficiary Records includes records from the following SSA systems of records: (1) Master Beneficiary Record, System of Record Notice 60-0090 (last published at 71 F.R. 1826 (January 11, 2006)) and (2) Supplemental Security Income Record and Special Veterans Benefits, System of Record Notice 60-0103 (last published at 71 F.R. 1830 (January 11, 2006)).

³ For Supplemental Security Income (SSI) purposes, SSA's search included individuals who are in active disability status, in addition to individuals who have been determined to be blind.

8. On May 28, 2014, SSA transferred responsive data to the Department of Justice, in the form of a 9.75 gigabyte, fixed width, encrypted text file. The name of the results file transferred to the Department of Justice is texas.dojfinal.0527.⁴
9. The results of each sweep, as described in Ex. A, can be found in the following fixed positions in the data that SSA returned to the Department of Justice on May 28:
 - a. Position 741 contains results of the Stage 1, Step 3.1.1, Combination A match;
 - b. Position 742 contains results of the Stage 1, Step 3.1.2, Combination B match;
 - c. Position 743 contains results of the Stage 1, Step 3.1.2, Combination C match;
 - d. Position 744 contains results of the Stage 1, Step 3.1.2, Combination D match;
 - e. Position 745 contains results of the Stage 1, Step 3.1.2, Combination E match;
 - f. Position 746 contains results of the Stage 1, Step 3.1.2, Combination F match;
 - g. Position 747 contains results of the Stage 2, Step 3.2.1, Combination G match;
 - h. Position 748 contains results of the Stage 2, Step 3.2.2, Combination H match;
 - i. Position 749 contains results of the Stage 2, Step 3.2.2, Combination I match;
 - j. Position 756 contains results of the Stage 2, Step 3.2.2, nine-digit social security number match;
 - k. Position 750 contains results of the Stage 2, Step 3.2.3, Combination K to G match;
 - l. Position 751 contains results of the Stage 2, Step 3.2.3, Combination K match;
 - m. Position 752 contains results of the Stage 2, Step 3.2.3, Combination K to L match;
 - n. Position 753 contains results of the Stage 2, Step 3.2.4, Combination L to G match;
 - o. Position 755 contains results of the Stage 2, Step 3.2.4, Combination L match;

⁴ SSA transferred identical results files to the Department of Justice on May 16, 2014 and May 23, 2014. These files could not be accessed by Department of Justice staff because of problems unencrypting the file.

- p. Position 754 contains results of the Stage 2, Step 3.2.4, Combination L to K match;
 - q. Position 757 contains results of the Stage 3, Step 3.3.1, Combination F match;
 - r. Position 758 contains results of the Stage 3, Step 3.3.2, Combination G match;
 - s. Position 759 contains results of the Stage 3, Step 3.3.2, Combination I match;
 - t. Position 766 contains results of the Stage 3, Step 3.3.2, nine-digit social security number match;
 - u. Position 760 contains results of the Stage 3, Step 3.3.3, Combination K to G match;
 - v. Position 761 contains results of the Stage 2, Step 3.3.3, Combination K match;
 - w. Position 762 contains results of the Stage 2, Step 3.3.3, Combination K to L match;
 - x. Position 763 contains results of the Stage 3, Step 3.3.4, Combination L to G match;
 - y. Position 765 contains results of the Stage 3, Step 3.3.4, Combination L match;
 - z. Position 764 contains results of the Stage 3, Step 3.3.4, Combination L to K match;
 - aa. Position 767 contains results of the Stage 4, Step 3.4.1, Sweep 1 match;
 - bb. Position 768 contains results of the Stage 4, Step 3.4.2, Sweep 2 match;
 - cc. Position 769 contains results of the Stage 4, Step 3.4.3, Sweep 3 match; and
 - dd. Position 770 contains results of the Stage 4, Step 3.4.4, Sweep 4 match.
10. For each of these positions, “1” indicates that the respective combination from one SSA record matched; “2” indicates that two or more SSA records matched on the respective combination; “0” indicates that there was no match; and a blank space indicates that no match was attempted.
11. When SSA initially ran the data comparison process, the results in Positions 750, 751, and 752 were grouped together rather than reported separately, as were the results in Positions 753, 754, and 755; Positions 760, 761, and 762; and Positions 763, 764, and

765. SSA did not produce those combined results to DOJ. Instead, at DOJ's request, SSA separated those results into separate positions, as reflected in the May 28 file. Likewise, when SSA initially ran the data comparison results for the Texas Stage 4 matches, only those SSA records that had not been matched on any of the Stage 1 to 3 matches were included in the eligible-for-matching population. SSA did not produce those results to DOJ. Instead, at DOJ's request, SSA re-ran the Texas Stage 4 matches, comparing the voter registration data to all eligible SSA records nationwide (as described in paragraph 6 above), regardless of whether they had been matched on any of the Stage 1 to 3 sweeps.

12. No other matching comparisons between the Texas data and data from the SSA were undertaken beyond those set forth above, and in accordance with the steps provided in Ex. A.

I declare under penalty of perjury that the foregoing is true and correct.

Executed on May 29, 2014.


Steve Strausler

Exhibit A

United States v. Texas: Federal Agency Algorithm Summary

This document summarizes the database matches that both the United States and all other plaintiffs and the State of Texas have requested from Federal Agencies as part of the *Veasey v. Perry/United States v. Texas* litigation (S.D. Tex).

The matching process proceeds in three parts, which are explained in detail below. *First*, databases are prepared and standardized. *Second*, identifier values are constructed by combining multiple individual fields. *Third*, the United States' one-to-many matches and the State of Texas's many-to-many matches are conducted between databases.

PART I: DATABASE PREPARATION

Stage 1: Extraction of Available Data from Federal Identification & Disability Databases

- Step 1.1.1:** Extract complete name into separate first name, middle name, and last name fields.
- Step 1.1.2:** Extract date of birth.
- Step 1.1.3:** Extract gender.
- Step 1.1.4:** Extract residential address and mailing address.
- Step 1.1.5:** Extract social security number.
- Step 1.1.6:** Extract Texas driver license number (only if present in Federal database).

Stage 2: Separate Valid Identification and Disability Records

- Step 1.2.1:** Remove records from identification database extracts that indicate that an ID has been revoked or has expired more than 60 days before the date of the TEAM database snapshot (which is January 15, 2014).

Exhibit A

- Step 1.2.2:** Remove records from disability database extracts that do not indicate current disability status or indicate a Veterans Administration disability rating of less than 50%.

Stage 3: Diagnostics

- Step 1.3.1:** Report the frequency of missing values for each field.
- Step 1.3.2:** Report the frequencies of invalid Social Security numbers, such as 111111111 and 123456789.
- Step 1.3.3:** Report the frequencies of likely invalid dates of birth, such as January 1, 1901 and November 11, 1911.

Stage 4: Standardize Last Name

- Step 1.4.1:** Remove last name suffixes that are contained within the last name field, rather than a distinct suffix field. *E.g.*, <Smith Jr.> becomes <Smith>.
- Step 1.4.2:** For last names containing hyphens, populate separate last name fields for all parts of the last name. *E.g.*, the last name <Smith-Jones> would have the value <Smith> entered into a LastName1 field and the value <Jones> entered into a LastName2 field.
- Step 1.4.3:** Remove spaces, hyphens, periods, and apostrophes from all last name fields and convert all letters to uppercase. *E.g.*, <O'Connor> becomes <OCONNOR> and <Smith-Jones> becomes <SMITHJONES>.
- Step 1.4.4:** Code all missing values as blank fields.

Exhibit A

Stage 5: Standardize First Name and Middle Name

- Step 1.5.1:** Remove spaces, hyphens, periods, and apostrophes from the first name field and convert all letters to uppercase. *E.g.*, <Jean-Paul> becomes <JEANPAUL>.
- Step 1.5.2:** Parse the first letter of the middle name (if available) and use it to populate a middle initial field. *E.g.*, <John> would yield <J>.¹
- Step 1.5.3:** Code all missing values as blank fields.

Stage 6: Standardize Date of Birth

- Step 1.6.1:** Convert the date of birth to an eight-digit string of MMDDYYYY.
- Step 1.6.2:** Code all missing values as blank fields.

Stage 7: Standardize Gender

- Step 1.7.1:** Code gender as a string of 1 for females and 0 for males.
- Step 1.7.2:** Fill missing gender values using the most common gender value for the first name associated with a record. *E.g.*, if 99% of records with first name <JOHN> are listed as male, assign the male identifier to all records with first name <JOHN> and no listed gender.
- Step 1.7.3:** If missing values remain, code all missing values as blank fields.

¹ The U.S. Department of State does not maintain a separate field for middle names in its database of U.S. Passport and Passport Card holders. Instead, both first and middle name may be stored in the first name field. For this database, the following rule will be applied: treat the first word in the first name field as the first name, and treat the first letter following the first space as the middle initial.

Exhibit A

Stage 8: Standardize Address

- Step 1.8.1:** Convert the residential ZIP code to a string if it is stored as a numeric field.
- Step 1.8.2:** Where the residential address ZIP code is blank, populate that field with the value in the mailing address ZIP code field, if available.²
- Step 1.8.3:** Truncate the residential ZIP code field to the first five digits. *E.g.*, <77777-1234> becomes <77777>.
- Step 1.8.4:** Where the residential address field is blank, populate that field with the value in the mailing address field, if available.
- Step 1.8.5:** Where address field containing street address begins with a street number, isolate the street number. *E.g.*, <123 Main Street> becomes <123>.
- Step 1.8.6:** Where the address field begins with recognized strings indicating a mail box, eliminate strings to isolate the box number. *E.g.*, <PO Box 444> becomes <444>.
- Step 1.8.7:** If missing values remain, code all missing values as blank fields.

Stage 9: Standardize Social Security Number

- Step 1.9.1:** Convert the social security number to a string if it is stored as a numeric field.
- Step 1.9.2:** Using full social security number, check for invalid SSNs. In the case of invalid SSNs, code as missing. *E.g.*, <123456789> becomes <>.

² For purposes of this database matching protocol, the only address fields utilized with respect to data regarding U.S. Passports and U.S. Passport Cards are those regarding mailing addresses.

Exhibit A

Step 1.9.3: Extract the last four digits of full social security number as a four-character string and use them to populate a separate SSN4 field.

Step 1.9.4: Code all missing values as blank fields.

PART II: DATABASE PREPARATION

Stage 1: Construct Primary Identifier Variables for United States' One-to-Many Sweeps

Step 2.1.1: Create Combination A: First Name + Last Name + Gender + DOB + Residential ZIP + Residential Street Number. *E.g.*, the separate fields <JEAN>, <SMITH>, <0>, <01011950>, <77777>, and <123> are combined to a single field <JEANSMITH00101195077777123>.³

Step 2.1.2: Create Combination B: Last Name + Gender + DOB + Residential ZIP + Residential Street Number.

Step 2.1.3: Create Combination C: Gender + Date of Birth + Residential ZIP + Residential Street Number.

Step 2.1.4: Create Combination D: First Name + Last Name + Date of Birth + Residential ZIP + Residential Street Number.

Step 2.1.5: Create Combination E: First Name + Last Name + Gender + Residential ZIP + Residential Street Number.

Step 2.1.6: Create Combination F: First Name + Last Name + Gender + DOB.

Step 2.1.7: Create Combination M: Texas Driver License Number (where available).

³ For the U.S. Department of State only, the name portion of any combination is truncated if it is more than 32 characters long.

Exhibit A

Stage 2: Construct Secondary Identifier Variables for United States' One-to-Many Sweeps

- Step 2.2.1:** Create Combination G: First Name + Middle Initial + Last Name + Date of Birth.⁴
- Step 2.2.2:** Create Combination H: SSN4 + Date of Birth + Residential ZIP.
- Step 2.2.3:** Create Combination I: SSN4 + First Name + Last Name + Date of Birth.
- Step 2.2.4:** Create Combination K: First Name + Last Name 1 + Middle Initial + Date of Birth.
- Step 2.2.5:** Create Combination L: First Name + Last Name 2 + Middle Initial + Date of Birth.
- Step 2.2.6:** Full Social Security Number.

Stage 3: Construct Identifiers Used Only For Texas's Many-to-Many Sweeps

- Step 2.3.1:** Create Combination for Texas's Sweep 1: SSN4 + Last Name + DOB.
- Step 2.3.2:** Create Combination for Texas's Sweep 3: First Name + Last Name + DOB
- Step 2.3.3:** Create Combination for Texas's Sweep 4: First Name + Middle Initial + Last Name + DOB

Note: Combinations for Texas's Sweeps 1 and 3 do not already exist as pre-made fields in the TEAM database extract but instead must be created from the underlying TEAM database fields, in addition to being constructed on the Federal database side. Texas's Sweep 4 is equivalent to

⁴ Only for the State Department, create three further variations of Combination G created using the State Department's "LFMName" field which contains Last, First, and Middle Names, in that order, truncated to a maximum length of 32 characters. Combination G1 is DOB + LFMName; Combination G2 is DOB + First two words of LFMName; and Combination G3 is DOB + First two words of LFMName + First character of third word of LFMName.

Exhibit A

the combination for the United States' Combination G. Texas's Sweep 2 is on full 9 social security number.⁵

Stage 4: Establish Identifier Uniqueness For Combinations A - L

Step 2.4.1: Generate a field that establishes the uniqueness of each identifier variable.

For federal databases, for each combination A-L, generate a field that establishes uniqueness among only Texas records and a field that establishes uniqueness among nationwide records. *E.g.*, if only one record has the string <JEANSMITH01011950012377777> for Combination A, populate the uniqueness field for Combination A for that record as <1>. If four records have the string <JOHNSMITHA0101950> for Combination G, populate the uniqueness field for Combination G for each of those records as <2>, which indicates any number greater than one.

⁵ For purposes of matching to U.S. Department of State Passport and Passport Card holder data, Texas's Sweep 3 and Sweep 4 are as follows: Sweep 3: Last Name + First Name (restricted to 32 characters) + DOB. Sweep 4: Last Name + First Name + Middle Initial (restricted to 32 characters) + DOB.

Exhibit A

PART III: MATCH DATABASES

Stage 1: United States' Primary One-to-Many Matching Sweeps

- Step 3.1.1:** For each case in which Combination A is unique in the TEAM database, match Combination A against Combination A in the identifier or disability database. For federal databases, use only the subset of records with Texas addresses in the identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination A output field. Where there is a match, indicate the uniqueness of Combination A in the identifier or disability database in the Combination A output field (*e.g.*, in cases where there is one matching record in the Federal database, <1> should be inserted into the Combination A output field, while a <2> should be inserted into the Combination A output field if the TEAM record matched 2 or more records in the Federal database).
- Step 3.1.2:** Use the procedure in Step 3.1.1 to match Combination B, Combination C, Combination D, Combination E, and Combination F in the TEAM database against the equivalent combination field in the identifier or disability database.
- Step 3.1.3:** Use the procedure in Step 3.1.1 to match Combination M in the TEAM database against the equivalent combination field in the identifier databases produced by the State of Texas.

Exhibit A

Stage 2: United States' Secondary One-to-Many Matching Sweeps

Step 3.2.1: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), and where Combination G is unique in the TEAM database, match Combination G against Combination G in the identifier or disability database. For federal databases, use only the subset of records with Texas addresses in the identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination G output field. Where there is a match, indicate the uniqueness of Combination G in the identifier or disability database in the Combination G output field (*e.g.*, <1> if a unique match and <2> if matched to more than one record).

Step 3.2.2: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination H, Combination I, and complete social security number⁶ in the TEAM database against the equivalent combination/field in the identifier or disability database.

Step 3.2.3: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination K against Combination G, Combination K, and Combination L in the identifier or disability database.

⁶ The full social security number is not created as a separate "combination" as it is its own field stored within the TEAM database under the field name "ssn".

Exhibit A

Step 3.2.4: For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination L against Combination G, Combination K, and Combination L in the identifier or disability database.

Stage 3: United States' Nationwide Federal Sweeps

Step 3.3.1: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, and where Combination F is unique, match Combination F against Combination F in the nationwide identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination F nationwide output field. Where there is a match, indicate the uniqueness of Combination F in the identifier or disability database in the Combination F nationwide output field (*e.g.*, <1> if a unique match and <2> if matched to more than one record).

Step 3.3.2: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination G, Combination I, and full social security number in the TEAM database against the equivalent combination/field in the nationwide identifier or disability database.

Step 3.3.3: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination

Exhibit A

K against Combination G, Combination K, and Combination L in the nationwide identifier or disability database.

Step 3.3.4: For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination L against Combination G, Combination K, and Combination L in the nationwide identifier or disability database.⁷

Stage 4: Texas' Many-to-Many Nationwide Sweeps

Step 3.4.1 Regardless of whether the combination for Sweep 1 is unique in the TEAM database, match against the equivalent combination in a nationwide search of the Federal database.

⁷ Step 3.3.5 for the State Department only: Match the following Combination G variations from applicable State Department records, first to include only the subset of records with Texas addresses, and then to include all applicable U.S. Passport and Passport Card records nationwide (*e.g.*, without Texas addresses), against the following fields from the TEAM database:

- Combination G1 to DOB + Last_fix + First_fix + Middle_name from the TEAM database;
- Combination G2 to DOB + Last_fix + First_fix from the TEAM database;
- Combination G2 to DOB + Last_fix + First word of First_name from the TEAM database;
- Combination G3 to DOB + Last_fix + First_fix + Middle_Initial from the TEAM database;
- Combination G3 to DOB + Last_fix + First word of First_name + Middle_Initial from the TEAM database; and
- Combination G3 to DOB + Last_fix + First word of First_name + First character of Second word of First_fix from the TEAM database.

Attempt matches for all TEAM records, regardless of whether they matched in any prior sweeps. Indicate <1> if a unique match and <2> if matched to more than one record.

Exhibit A

Step 3.4.2 Regardless of whether full 9 social security number is unique in the TEAM database, for Sweep 2, match against the equivalent field in a nationwide search of the Federal database.

Step 3.4.3 Regardless of whether the combination for Sweep 3 is unique in the TEAM database, match against the equivalent combination in a nationwide search of the Federal database.

Step 3.4.4 Regardless of whether the combination for Sweep 4 (Combination G) is unique in the TEAM database, match against the equivalent combination in a nationwide search of the Federal database.

Note: For each of the Texas many-to-many sweeps:

- Indicate <1> if any TEAM combination matches a single combination in the Federal database
- Indicate <2> if any TEAM combination matches more than one record in the Federal database.
- Indicate <0> if no match is achieved.

Examples:

- If there are two TEAM records that have identical versions of the combination for Sweep 1, and there is one record in the Federal database that matches on this combination, both of the underlying TEAM records will have a matching output of <1> for Sweep 1.
- If there are three TEAM records that have identical versions of the combination for Sweep 3, and there are five records in the Federal database that match on that combination, the three TEAM records will each have a matching output of <2> for Sweep 3.

Exhibit B

Social Security Administration

Step 1.3 Diagnostics on SSA Data

Step 1.3.1:

Missing first name – 0%
Missing middle name – 20.944%
Missing last name – 0%
Missing date of birth – 0%
Missing gender – 0%
Missing residential address – 63.938%
Missing mailing address – 0%
Missing SSN – 0.0264%

Step 1.3.2:

Invalid SSN – 0%

Step 1.3.3

Invalid date of birth – 0%